FINAL TECHNICAL REPORT

THE NATIONAL CENTER FOR COLLABORATION IN MEDICAL MODELING AND SIMULATION

MAY 2005

C. Donald Combs
EASTERN VIRGINIA MEDICAL SCHOOL
NORFOLK, VIRGINIA

Prepared For The

NAVAL AIR SYSTEMS COMMAND
NAVAL AIR WARFARE CENTER TRAINING SYSTEMS DIVISION
12350 Research Parkway
Orlando, FL 32826-3275

Approved for public release; distribution is unlimited.

William J. Wasilenko
Associate Dean for Research
Eastern Virginia Medical School

C. Donald Combs
Principal Investigator
Eastern Virginia Medical School

GOVERNMENT RIGHTS IN DATA STATEMENT

Reproduction of this publication in whole
or in part is permitted for any purpose
of the United States Government.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
20-05-2005	FINAL TECHNICAL REPORT	23-09-2003 - 31-01-2005

4. TITLE AND SUBTITLE
The National Center for Collaboration in Medical Modeling and Simulation

5a. CONTRACT NUMBER
N61339-03-C-0157

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
CDRL-A002

6. AUTHOR(S)
Combs, C. Donald

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
William J. Wasilenko, Ph.D., Associate Dean for Research
Eastern Virginia Medical School
358 Mowbray Arch
Norfolk, VA 23507-2219

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Air Warfare Center Training Systems Division
12350 Research Parkway
Orlando, FL 32826-3275

10. SPONSOR/MONITOR'S ACRONYM(S)
NAVAIR TSD

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

20050606 027

14. ABSTRACT
This project demonstrates the objective value of medical simulations as training tools for use by military medical personnel in training for tasks that are relevant to the effective and efficient medical care of military pesonnel in combat settings as well as in CONUS hospitals and clinics operated by the military. In addition, the project pursues an integration effort to provide a coherent set of medical simulations and the related medical education/training curricula in which these simulations will be used. Finally, the project provides an architecture and a specific example of a regional model and associated simulations for use in preparing a CONUS region for its medical response to natural or man-made events involving mass destruction.

15. SUBJECT TERMS
Medical modeling, medical simulation, surgical simulation, augmented reality

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	UU	192	C. Donald Colmbs, Ph.D.
U	U	U			19b. TELEPONE NUMBER *(Include area code)* 757-446-6090

ACKNOWLEDGEMENTS

We appreciate the research support and assistance provided by the researchers and staff of the Virginia Modeling, Analysis and Simulation Center (VMASC) of Old Dominion University led by R. Bowen Loftin, Ph.D. We are grateful to the staff of the National Capital Area Medical Simulation Center of the Uniformed Services University of the Health Sciences. In addition, we acknowledge the research assistance of the staff of the Naval Medical Center-Portsmouth. Finally, we thank the medical students and medical residents of the Eastern Virginia Medical School and the Uniformed Services University of the Health Sciences and graduate students of Old Dominion University without whom this research would not have been possible.

EXECUTIVE SUMMARY

PROBLEM/BACKGROUND

The quality and cost of medical care are among the dominant issues of the twenty-first century. The availability and preparation of physicians and other medical care providers for support of military operations is threatened by the relative lack of opportunities to practice battlefield medicine, of support for civilian and military response to bio-terrorism and other weapons of mass destruction, and by the need to improve the quality of medical care available to the military services.

OBJECTIVE

This project demonstrates the objective value of medical simulations as training tools for use by military medical personnel in training for tasks that are relevant to the effective and efficient medical care of military personnel in combat settings as well as in CONUS hospitals and clinics operated by the military. In addition, the project pursues an integration effort to provide a coherent set of medical simulations and the related medical education/training curricula in which these simulations will be used. Finally the project provides an architecture and a specific example for a regional model and associated simulations for use in preparing a CONUS region for its medical response to natural or man-made events involving mass destruction.

APPROACH/ORGANIZATION OF THE REPORT

Approaches utilized in the project research included validating existing medical simulations, developing new medical simulations, creating databases, examining curricula and developing an architecture and a specific example for the regional mass casualty simulation, among others.

FINDINGS/RESULTS

Results of the analysis revealed that some medical simulations, albeit low-tech in nature, may be more effective than other high-tech simulations, that new simulations under development may be useful in medical and surgical education, that there has been up to now no single repository of medical modeling and simulation data, that simulation has an appropriate place in medical education curricula and that the effects of a mass casualty event on a region can be effectively simulated.

CONCLUSIONS

Additional multi-center research with a larger "n" is necessary to more conclusively validate medical simulations. Additional research is necessary to further

EXECUTIVE SUMMARY (Continued)

develop those medical simulators under development under this project. Incorporating simulations in medical education curricula is difficult in a time-limited environment with multiple administrative entities involved. It has been possible to develop a comprehensive medical modeling and simulation database. Additional research is necessary to more comprehensively model a mass casualty incident in a multi-jurisdictional region such as Hampton Roads, Virginia.

RECOMMENDATIONS

Future simulator validation studies need to incorporate additional subjects among multiple research centers if necessary. More realistic timetables for the development of new simulations should be adopted that reflect the complexity of this task. Medical education curricular redesign should incorporate validated medical modeling and simulation tools. The Medical Modeling and Simulation Database should be updated on a regular basis to keep it useful for researchers in the field. Additional research on mass casualty simulations needs to be undertaken as this is a very complex undertaking.

FINAL TECHNICAL REPORT

TABLE OF CONTENTS

FINAL TECHNICAL REPORT

PROBLEM/BACKGROUND

SIGNIFICANCE

The year 1999 marked the seventieth anniversary of the development of the first flight simulator by Edwin Link. This device, derived from a carnival ride, successfully trained thousands of American aviators during World War II and contributed immeasurably to the success of the Allied military effort. Now, at the onset of the twenty-first century, flight simulation has reached such a level of maturity that many commercial pilots fly their first plane load of passengers in a new aircraft with only simulation experience in "that" aircraft. It is time to explore what the field of modeling and simulation can offer to the likely most dominant enterprise of the next century – the evolution of the practice of medicine. In its recent report, *To Err is Human*, the Institute of Medicine highlighted the use of lessons learned in flight simulation as a means to better train current and future medical personnel in one of its key recommendations:

> *The Committee believes that health care organizations should establish team training programs for personnel in critical care areas (e.g., the emergency department, intensive care unit, operating room) using proven methods such as the crew resource management techniques employed in aviation, including simulation.* *

From this report and similar sources, a strong case can be made for a national effort to exploit the national investment, predominantly in the military, in modeling and simulation for the training of those delivering medical care and for the provision of assistance during the delivery of medical services. Such an effort will leverage seventy years of experience in modeling and simulation and serve to decrease the cost of and increase the quality and availability of medical care.

MILITARY RELEVANCE

In the event of a major military operation outside of the continental U.S., reservists called up for active duty supplement regular military medical personnel. This is especially true during wartime. For example, a large fraction of medical personnel deployed to the Persian Gulf during Operation Desert Shield and Operation Desert Storm were reservists. These reservists are unlikely, in general, to have current experience with the trauma wounds and exposure to chemical or biological agents that could be sustained by combat and combat support personnel.

The quality and cost of medical care are among the dominant issues of the twenty-first century. The availability and preparation of physicians and other medical care providers for support of military operations is threatened by the lack of opportunities to practice battlefield medicine, of support for civilian and military response to bio-terrorism and other weapons of mass destruction, and by the need to improve the quality of medical care available to the military services.

* Committee on Quality of Health Care in America, Institute of Medicine. *To Err is Human: Building a Safer Health System*. Linda T. Kohn, Janet M. Corrigan, and Molla S. Donaldson, Editors. Washington, D.C.: National Academy Press, 1999, p. 149.

OBJECTIVE

The research conducted by the National Center for Collaboration in Medical Modeling and Simulation through this project demonstrates the objective value of medical simulations as training tools for use by military medical personnel in training for tasks that are relevant to the effective and efficient medical care of military personnel in combat settings as well as in CONUS hospitals and clinics operated by the military. In addition, the project pursues an integration effort to provide a coherent set of medical simulations and the related medical education/training curricula in which these simulations will be used. Finally, the project provides an architecture and a specific example for a regional model and associated simulations for use in preparing a CONUS report for its medical response to natural or man-made events involving mass destruction. Such events inevitably rely upon military medical resources. The model and simulations include the military component of the medical response.

The objectives for the research undertaken fall into four broad categories: simulator validation, technology development, simulator and curricular integration and regional medical response simulations. In simulator validation, the key objective was to conduct validation studies of various medical simulators; in technology development, the key objective was to begin the selective development of medical and surgical simulators; in simulator and curricular integration, the key objective was to conduct surveys of medical curricula to determine where simulation is the most appropriate training aid; and, in regional medical response simulations, the key objective was to develop a conceptual model of the simulation based on system-level requirements.

APPROACH/ORGANIZATION OF THE REPORT

The Technical Report is organized so that its sections directly correspond with the contract requirements as expressed in the major work task subheadings in the underlying contract between NAVAIR ORLANDO TSD and Eastern Virginia Medical School to undertake the National Center for Collaboration in Medical Modeling and Simulation research portfolio. Thus, the Technical Report section headings that follow are Task 1 – Simulator Validation, Task 2 – Technology Development, Task 3 – Simulators and Curricular Integration and Task 4 – Regional Medical Response Simulations. Under each major work task subheading there is a further breakdown into distinct minor work tasks that have been indicated with an alphabetical system ((a), (b), (c)) which, in turn, also corresponds with the contract requirements in the underlying contract. This approach to organizing the Technical Report allows the reader to easily locate research results in the Technical Report that are related to specific contract requirements.

TASK 1

SIMULATOR VALIDATION

FINAL TECHNICAL REPORT

Task 1 - Simulator Validation.

The contractor shall:

(a) Conduct validation experiments of the catheter insertion simulator using medical, physician's assistant, and psychology students.

Phase 1 of the catheter insertion simulator validation experiment was completed. Experiments using EVMS medical students (Appendix A.1), EVMS physician assistant students (Appendix A.2), and ODU psychology students (Appendix A.3) were executed. Each experiment resulted in a presentation at a professional meeting and/or publication. The specifics are noted in each appendix.

(b) Conduct validation experiments of the colonoscopy simulator using surgical and psychology students.

Phase 1 of the colonoscopy simulator validation experiment in still in progress. Data collection for the experiment using EVMS surgical residents is in progress. Data collection for the experiment using ODU psychology students has been completed and the results are now being analyzed.

(c) Conduct validation experiments of the high performance skill medical simulators using surgical and psychology students.

The high performance skill medical simulators being developed under this research program are the wound debridement simulator and the augmented standardized patient simulator. These simulators are both currently in prototype stage. Validation experiments will commence after the designs of the high performance skill medical simulators are finalized and a final prototype is developed.

(d) Provide support for validation experiments of an arthroscopy simulator using surgical staff and psychology students.

Phase 1 of the arthroscopy simulator validation experiment in still in progress. The scoring system utilized was developed here at the Naval Medical Center-Portsmouth. The scoring system is divided into areas of the knee that must be visualized during surgery. Point values were set for each area to correspond to its level of importance. Once the scope is completed a total score of 75 points is possible. An additional 25 points is awarded based on the time it takes to complete the arthroscopy. In addition, any missed items would result in a time penalty which decreases the time points that can be earned. This system is designed to evaluate a surgeon both on his ability to view all required areas of the knee and his ability to do this in a timely manner. Surgeons need to be both thorough and efficient in the operating room. Based on a meeting with all staff members, it was felt that the minimum competency for a surgeon would be to complete a diagnostic arthroscopy in 7 minutes or less.

The scoring system was then taken through a validation study which included both residents and staff surgeons that perform arthroscopy as part of their practice. This was a blinded prospective randomized study. Using an ANOVA test with a follow-up Tukey-Kramer test, a significant difference was noted between residents at different levels of training and staff surgeons $p<0.001$. This verified that the scoring system could differentiate between novice, intermediate and experts (staff surgeons) in diagnostic arthroscopy.

The second phase of this study is designed to determine if there is transference from the simulator to the operating room. The subjects initially perform a diagnostic knee arthroscopy on a patient. During this procedure they are scored to determine their baseline skill level. They are then shown a demonstration of the simulator and then scored while doing a baseline simulated arthroscopy. The subjects are then instructed on the arthroscopy simulator. They then work on the simulator until they can achieve minimum competency. The individuals are then scored again on the simulator to verify they can meet the minimum competency for arthroscopy. This is a total score of 85 points; this corresponds to doing a complete diagnostic arthroscopy in 10 minutes or less. Once this level is achieved the subject is then rescored in the operating room while performing a diagnostic arthroscopy on a patient. The overall process takes approximately 4 weeks to complete. This occurs because the surgical residents must work on the simulator after hours and remain under the mandatory 80 hour work week. Due to schedule restrictions the residents are also only available to perform the clinical portion of this study on Fridays.

We have enrolled 13 subjects to date, but 4 of the senior level residents tested at the minimum competency level on the initial arthroscopy so they will not be included in the study. Over the next 5 months we will enroll another 5 subjects that are all junior in their training. We also plan on enrolling another 8 subjects in the next academic year for a total of 22 subjects. It is anticipated that the study will be completed in June 2006. Data collection for the experiment using ODU psychology students has been completed (Appendix A.4).

(e) Establish research protocol(s) for data collection.

Research protocols were established for a component of the catheter insertion simulator validation experiment with research staff at the National Capital Area Medical Simulation Center of the Uniformed Services University of the Health Sciences (USUHS). An experiment was carried out and USUHS has provided a research paper summarizing those results (Appendix A.5).

(f) Perform literature review of research on high performance skill acquisition.

The literature on high performance skill acquisition was reviewed by human factors specialists prior to the development of the wound debridement and augmented standardized patient high performance skill medical simulators. Also, in support of the wound debridement simulation development effort, an extensive literature search has been conducted in various areas of tissue modeling including the review of over 150 scholarly works. In addition, the Medical Modeling and Simulation Database (MMSD) has been created. Please refer to Task 1.h (see below).

(g) Provide human factors input into the development of high performance skill medical simulators.

Human factors requirements and specifications were provided by human factors psychologists for the development of the wound debridement and augmented standardized patient high performance skill medical simulators (see Task 2).

(h) Perform literature reviews of Virtual Reality (VR) medical simulation research.

A literature review of Virtual Reality (VR) medical simulation research was initiated and is still in progress. The Medical Modeling and Simulation Database (MMSD) has been created. The MMSD consists of two web-based, searchable compilations: one, the Research Database, that contains bibliographic information on published articles and abstracts (where available) and a second, the Companies and Projects Database, that maintains contact information for research centers, development and application programs, journals and conferences. The MMSD has been developed to increase awareness of the breadth of the medical modeling domain and to provide a means for fostering collaboration by bringing like-minded organizations and researchers into more frequent contact with each other, thus speeding advancement of the medical modeling and simulation domain.

(i) Organize professional meetings to address the current state of modeling and simulation methods and technology and its applicability to medicine.

Plans to organize a professional meeting to address the current state of modeling and simulation methods and technology and its applicability to medicine have been discussed. Work on a professional meeting will proceed pending the availability of funds in subsequent years.

(j) Conduct a research program to study the practice of medicine/surgery (both research and application issues).

The practice of medicine and surgery was studied. Because of the changing nature of the practice of medicine and surgery, the paradigm of medical education must also change, requiring the use of alternative instructional methodologies. Simulation training is a viable alternative that allows the learner to obtain experience and skill prior to interacting with patients in vivo. The increasing focus on issues of patient safety, health care costs and liability in medical and surgical practice provides further impetus for these changes. Many medical education institutions provide some level of simulated training to their medical and surgical community. However, surgical training programs, overall, lack substantive training utilizing simulation, especially in advanced minimally invasive techniques, which are rapidly changing the nature of the practice of surgery. Minimally invasive surgery is surgery performed utilizing small incisions, requiring surgeons to rely on a camera and video monitor while navigating instruments within a tiny field of view. The benefits of minimally invasive surgery for the patient include smaller incisions that frequently result in decreased blood loss and less pain, a quicker period of recovery, and decreased healthcare costs.

Surgical residency programs must improve their training and education in this area in order to graduate surgeons who are proficient in this skill. Experts in minimally invasive surgical training believe that surgical residents must perform these procedures many more times than is currently recommended by the Residency Review Committee in Surgery. The challenge remains how to restructure an already crowded surgical resident curriculum to include training and education for new advancements in surgical techniques in surgical practice at a time when the clinical and formal educational activities of surgical residents have been restricted to 80 hours per week effective 7/1/03.

In response to these types of changes in surgical and medical practice, a comprehensive review of the third and fourth-year undergraduate medical education curriculum has been completed that identifies current simulation training offered and opportunities to include simulation training for General Surgery, Otolaryngology-Head and Neck Surgery, Urology, Obstetrics and Gynecology, and Emergency Medicine. Similar curriculum reviews have also been completed for the Surgical Assistant program curriculum, for Emergency Medicine resident physicians, General Surgery residents, for Pediatric residents, for Obstetrics and Gynecology residents, for Otolaryngology-Head and Neck Surgery residents, for Urology residents, and for Vascular Surgery fellows.

TASK 2

TECHNOLOGY DEVELOPMENT

Task 2 - Technology Development.

The contractor shall:

(a) Initiate and maintain an inventory of existing Department of Defense (DOD) medical modeling and simulation activities, current projects and points of contact.

An inventory of existing DOD medical modeling and simulation activities has been completed. The document "Patterns of Disease in the U.S. Military: Looking Back, Thinking Ahead" (Appendix B.2) is a modeling study of the profiles of diseases within and among the various United States military services. The study concludes, among other things, that two distinct disease patterns can be identified in the U.S. military, that of peacetime hospitalization and that of wartime hospitalization and peacetime outpatient illness. In addition, the study concludes that when combat status is held constant, the general disease patterns of specific populations vary somewhat across the military services and within subpopulations within each service. Also, the Medical Modeling and Simulation Database (MMSD) has been created. The MMSD consists of two web-based, searchable compilations: one, the Research Database, that contains bibliographic information on published articles and abstracts (where available) and a second, the Companies and Projects Database, that maintains contact information for research centers, development and application programs, journals and conferences. The MMSD has been developed to increase awareness of the breadth of the medical modeling domain and to provide a means for fostering collaboration by bringing like-minded organizations and researchers into more frequent contact with each other, thus speeding advancement of the medical modeling and simulation domain.

(b) Develop collaborative mechanisms for medical modeling and simulation activities nationwide.

Several activities conducted during the period of performance have supported the development of collaborative activities. Several members of the research team have attended and presented research papers at conferences nationally and internationally. Research team members have joined organizations such as the Society for Medical Simulation and have participated in initiatives with other researchers in the field such as the Advanced Initiatives in Medical Simulation (AIMS) initiative. In addition, the research team has begun correspondence and collaborative research with other groups nationally that are conducting complementary research, such as the National Capital Area Medical Simulation Center of the Uniformed Services University of the Health Sciences. In addition, the Medical Modeling and Simulation Database (MMSD), discussed above, was designed as a means to foster collaboration in the medical modeling and simulation field. A planned collaborative effort with the American College of Surgeons will make the MMSD more user-friendly and provide additional technical and marketing expertise for the project.

(c) Acquire a range of medical and surgical simulators for use in validation and training transfer studies.

In this year's contract, several pieces of equipment have been added using project funds.

The Mentice Virtual Arthroscopy knee simulator has been purchased for research use by orthopedic surgeons at the Naval Medical Center-Portsmouth in conjunction with EVMS and ODU researchers. These researchers are studying the use of the simulator in training orthopedic surgery residents on repairing the knee injuries that have become so common in troops serving in Afghanistan and Iraq.

A research study of surgical interns is underway utilizing a colonoscopy simulator purchased utilizing project funds. Training methodologies being compared in the study include performing colonoscopies in vivo versus utilizing the colonoscopy simulator.

A Beowulf cluster was purchased from Professional Service Super Computers (PSSC) by EVMS which is currently being used to develop and test algorithms for efficiently simulating pathogen antigenic variation and host immune response. This work is part of a larger research program on the effects of partial and cross-reactive immunity on pathogen antigenic diversity within host populations. The work focuses on infectious disease transmission system dynamics and will use the cluster to simulate the transmission of antigenically distinct pathogens through populations of hosts with individual dynamic immune responses. EVMS also plans to utilize the cluster for evaluating biostatistical sampling methodology, bioinformatics and exploring novel discrimination methods for clinical and scientific decision-making.

EVMS has partnered its expertise in epidemiology and virology with the Virginia Modeling, Analysis and Simulation Center's expertise in large-scale population simulation and submitted a proposal to the National Science Foundation for a competitive multi-scale modeling grant. If funded, the grant would allow EVMS to study pathogen transmission and evolution in a simulated host population of at least 50,000 autonomous "intelligent-agents" with dynamic immune response. The simulations would be run on VMASC's large Beowulf cluster. The proposed pathogen model for the project is an enteric virus with a single-stranded RNA genome that would have a high rate of mutation and a great deal of antigenic diversity.

A Mentice MIST-VR minimally invasive surgical trainer with suturing module and camera navigation module was purchased from Medical Education Technologies, Inc. for use by the EVMS Department of Surgery for training and research on laparoscopic surgical procedures. Research is currently in the very early stages with this piece of equipment. Surgical interns are currently practicing basic laparoscopic skills with a pelvi-trainer simulator acquired this year.

A number of inanimate models have been acquired from companies such as Limbs and Things, Medisim, Nasco Arts and Crafts and others to support training and research activities in the EVMS Department of Surgery. A suturing workshop was developed for third year medical students in the surgery clerkship that will incorporate a research study on the training of suturing technique utilizing such models. PGY-2 surgery residents will receive advanced laparoscopic training sessions utilizing simulated gallbladders to hone these skills.

Additional equipment has been purchased with Old Dominion University funds that will support surgical simulation efforts. Four haptics devices that can support and advance surgical simulation research efforts have been purchased. These devices are two Phantom 1 5/6 DOF, and two Phantom Omni haptics devices.

(d) Survey existing tissue models and simulator technologies and begin selective development of medical and surgical simulation.

Technology development has progressed along two thrusts. The first and principal thrust is the development of a simulator environment that would support medical training for wound debridement. In support of the wound debridement simulation development effort, an extensive literature search has been conducted in various areas of tissue modeling including the review of over 150 scholarly works. The second thrust resulted in the augmented standardized patient, a standardized patient augmented with computer simulation generated indications that they are incapable of producing.

Wound Debridement Simulation

Wound debridement refers to the removal of necrotic, devitalized, or contaminated tissue and/or foreign material to promote wound healing. One method to accomplish this procedure is by surgically cutting away the necrotic tissue. Typically, this procedure is taught using actual patients. Simulator-based training can ensure that a medical trainee is competent prior to performing a procedure on a genuine patient; it can extend the range of experience with different types of wounds; introduce different complications; and it can provide instruction and timely performance feedback without requiring the presence of more experienced medical personnel. Through integration with medical curricula our wound debridement trainer also aims to increase the number of trained medical personnel.

The wound debridement trainer comprises a multimedia overview (Figure 1); a surgical simulator (tutorial component); and an assessment component. The initial training module provides an introduction and overview of the debridement process including information on procedures, instruments and wound types. This multimedia (web-based) component can launch a tutorial in the (second) simulator component or can run independently of the simulator since it requires only a web browser.

The surgical simulator provides the opportunity to prepare the site by scrubbing and rinsing the wound, to remove foreign objects, and to perform debridement on a virtual patient. Trainees must also learn to avoid cutting critical structures including blood vessels and nerves. Our simulator aims to deliver highly realistic graphics, real-time performance feedback, and haptic response to the trainee.

Simulation requirements currently under research and development include the following:

- Tissue deformation modeling using both mass-spring models and accelerated finite element modeling;
- Collision detection and cutting;
- Soft-body haptics rendering for two devices in the same space (one for each hand); and

- Realistic graphics rendering and visualization of anatomical data from the National Library of Medicine's Visible Human project to ensure correct anatomies and wounds.

Evaluation of trainee performance is another important part of the simulator. The tutorial mode records interactions and provides performance-based feedback during training and summative feedback at the end. The assessment mode provides only summative feedback upon completion of the whole procedure.

A rapid prototype of the surgical trainer has been developed and comprises an interactive mass-spring model of wound tissue. The system allows the trainee to clean debris from the wound using a syringe and saline (Figures 2, 3) and to remove glass shards with forceps (Figure 4). A Phantom device from SenAble Technologies, Inc. is used for the tracking and interaction of one hand, and the environment is viewed on a Reachin Technologies, AB, mirrored display.

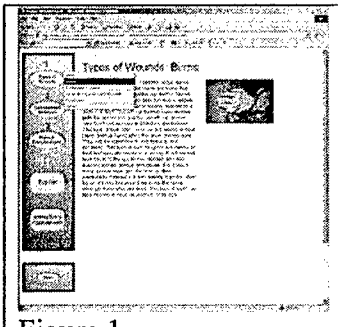


Figure 1.


Figure 2.

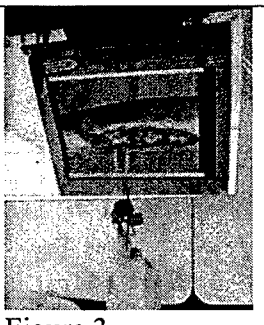


Figure 3.

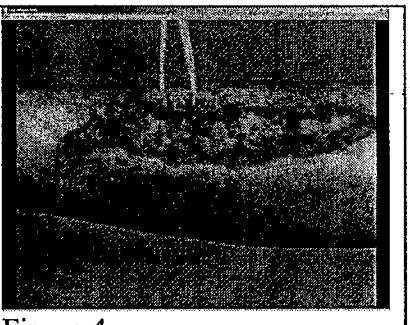


Figure 4.

Augmented Standardized Patient

In medical education, standardized patients are actors that are trained to present various medical indications to train medical students and other medical personnel in the art of making a diagnosis. Because standardized patients are typically healthy individuals, they have a limited ability to present symptoms for many common conditions. In order to extend the education and training capability, the augmented standardized patient (ASP) was conceived. In the ASP, the standardized patient is engaged and his/her symptoms are augmented with computer simulations to provide a broader range of learning opportunities for medical students. At the end of the period of performance, a working prototype has been constructed providing the ability to auscultate computer processed normal heart sounds at different locations on the ASP. One publication resulted during the period of performance and is included as an appendix to this report (See Appendix B.1).

TASK 3

SIMULATORS AND CURRICULUM INTEGRATION

Task 3 - Simulators and Curriculum Integration.

The contractor shall:

(a) Identify requirements for high performance skill medical simulators.

The literature on high performance skill acquisition was reviewed by human factors specialists prior to the development of the wound debridement and augmented standardized patient high performance skill medical simulators. In support of the wound debridement simulation development effort, an extensive literature search has been conducted in various areas of tissue modeling including the review of over 150 scholarly works. Simulation requirements currently under research and development include the following:

- Tissue deformation modeling using both mass-spring models and accelerated finite element modeling;
- Collision detection and cutting;
- Soft-body haptics rendering for two devices in the same space (one for each hand); and
- Realistic graphics rendering and visualization of anatomical data from the National Library of Medicine's Visible Human project to ensure correct anatomies and wounds.

For the augmented standardized patient simulator, a significant amount of research was conducted into sensor technology and capabilities, into three-dimensional positional tracking devices based on both magnetic and infrared technologies, into electronic stethoscope technology, into the processing of sound files and into those elements of the physical examination that needed to be replicated with the augmented standardized patient.

(b) Solicit and identify use cases that describe the intended use of the high performance skill medical simulators support for research, planning, and training.

For the wound debridement simulator, which is currently in the prototype stage, a use case has been developed for the removal of shards of glass from a wound that includes a training component. The wound debridement trainer comprises a multimedia overview; a surgical simulator (tutorial component); and an assessment component. The initial training module provides an introduction and overview of the debridement process including information on procedures, instruments and wound types. This multimedia (web-based) component can launch a tutorial in the (second) simulator component or can run independently of the simulator since it requires only a web browser.

The surgical simulator provides the opportunity to prepare the site by scrubbing and rinsing the wound, to remove foreign objects, and to perform debridement on a virtual patient. Trainees must also learn to avoid cutting critical structures including blood vessels and nerves. Our simulator aims to deliver highly realistic graphics, real-time performance feedback, and haptic response to the trainee.

For the augmented standardized patient project, the use case developed envisions the simulator to be utilized initially in medical education research for determining its ability to provide improved instruction in conducting a physical examination of a patient versus the currently utilized approach. A research project involving the Objective Structured Clinical Examination (OSCE) required of all fourth-year medical students utilizing the augmented standardized patient technology is currently being designed for implementation in the summer of 2005. The outcomes of this study will drive further use case development in the areas of planning and training for a variety of health professions disciplines.

In addition, a project aimed at studying the nature of training and observing surgical skills under hazardous conditions was initiated. Two procedures, tube thoracostomy and cricothyroidotomy, were studied with medical students and surgical residents, respectively. Each experiment resulted in a presentation at a professional meeting and/or publication (Appendices C.1 and C.2).

(c) Develop detailed training simulator requirements based on the identified use cases.

Detailed training simulator requirements based on the identified use cases for the wound debridement and augmented standardized patient high performance skill simulators previously discussed are currently under development.

(d) Select a minimally invasive surgical procedure and conduct research into haptic and visual training simulators that address this surgical procedure.

Surgical interns are practicing basic laparoscopic skills using the Pelvi-trainer. The interns have timed drills that occur every 2-3 months. A database of their results is in progress. In addition, later this year PGY-2 surgical residents will receive advanced laparoscopic training sessions using simulated gallbladders to hone these skills.

(e) Research both commercially available virtual reality systems and state-of-the-art virtual reality research (e.g. universities) to determine the best development strategies.

The Medical Modeling and Simulation Database (MMSD) has been created. The MMSD consists of two web-based, searchable compilations: one, the Research Database, that contains bibliographic information on published articles and abstracts (where available) and a second, the Companies and Projects Database, that maintains contact information for research centers, development and application programs, journals and conferences. The MMSD has been developed to increase awareness of the breadth of the medical modeling domain and to provide a means for fostering collaboration by bringing like-minded organizations and researchers into more frequent contact with each other, thus speeding advancement of the medical modeling and simulation domain (Appendices C.3 and C.4).

(f) Initiate collaborations with relevant institutions for joint research and development.

The research team has begun correspondence and collaborative research with other groups nationally that are conducting complementary research, such as the National Capital Area

Medical Simulation Center of the Uniformed Services University of the Health Sciences. Discussions are underway to initiate a collaborative effort with the American College of Surgeons that will make the Medical Modeling and Simulation Database (MMSD) more user-friendly and will provide additional technical and marketing expertise for the project.

(g) Develop a strategic plan and documentation guidelines for the construction of medical and surgical training simulators that meet the researched standards and application.

Through the process of developing the two high performance skill medical simulators currently under construction, the wound debridement and the augmented standardized patient simulators, a strategic plan and documentation guidelines for the construction of medical and surgical training simulators is under development and will likely include a "lessons learned" document that will be useful to other institutions who seek to undertake the construction of medical and surgical training simulators.

(h) Establish medical and surgical training simulator standards.

It is envisioned that the Medical Modeling and Simulation Database (MMSD), discussed earlier, will be a very influential vehicle in the creation of medical and surgical training simulator standards. The next phase in the development of the MMSD will involve the creation of a "Morningstar" type of heuristic that will rate medical and surgical training simulators in a heuristic schema that incorporates at least three factors: the quality of the research design, the clinical importance of the simulator (a function of frequency of the procedure simulated and its cost) and the potential for leading to improved clinical care. The underlying criteria to be developed in instituting this rating system will serve as *de facto* medical and surgical training simulator standards.

(i) Conduct surveys of medical curriculum reviews to determine where simulation is the most appropriate training aid.

A comprehensive review of the third and fourth-year undergraduate medical education curriculum has been completed that identifies current simulation training offered and opportunities to include simulation training for General Surgery, Otolaryngology-Head and Neck Surgery, Urology, Obstetrics and Gynecology, and Emergency Medicine. Similar curriculum reviews have also been completed for the Surgical Assistant program curriculum, for Emergency Medicine resident physicians, General Surgery residents, for Pediatric residents, for Obstetrics and Gynecology residents, for Otolaryngology-Head and Neck Surgery residents, for Urology residents, and for Vascular Surgery fellows.

(j) Identify requirements for and initiate development of a laboratory dedicated to simulation-based training of minimally invasive surgical techniques and medical procedures.

Space has been identified and preliminary architectural plans drawn up to develop a Minimally Invasive Surgical Simulation Laboratory at Eastern Virginia Medical School. Eastern Virginia Medical School staff visited similar laboratories at St. Peter's University Hospital in New Brunswick, NJ, Johns Hopkins University Hospital in Baltimore, MD and the Montefiore

Hospital Institute for Minimally Invasive Surgery in the Bronx, NY in the course of development of plans for a similar center at Eastern Virginia Medical School. The proposed laboratory will be electronically linked to an operating room at Sentara Norfolk General Hospital and to the Eastern Virginia Medical School Theresa A. Thomas Professional Skills Teaching and Assessment Center as well as to video communication links via internet, microwave and satellite technology to allow a variety of training opportunities.

A full service laboratory dedicated to simulation of minimally invasive surgical procedures needs a physical space large enough to house a conference room and dry and wet laboratory operating space. This allows for the didactic presentation and discussion of any procedure under consideration. The conference space by necessity requires network connectivity to the web. The ability to connect real time to the operating room in the hospital, other conference rooms on campus or other such laboratories is essential. Networking into larger teaching programs utilizing faculty and students in remote locations will be an important function of a regional surgical simulation training laboratory located at a major medical center.

A dry procedure area will allow participants to utilize a variety of mechanical and virtual skills trainers that are essential to performance of complex surgical procedures. Sufficient space is necessary to house an inventory of such units. Certain units may be grouped together to simulate a particular scenario or training need. Space, upkeep and the ability to obtain or develop such materials are critical. Duplication of individual devices allows multiple users simultaneously.

A wet procedure area would be required so the trainee could perform the skill in non-survival large animal models as a prelude to human surgery until such high fidelity virtual models are available. Such a facility needs to mimic a real operating room in terms of the laparoscopic equipment so that the simulated procedure is "real" to the practitioner.

Appropriate locker facilities for guest students and faculty, lounge areas and supervisory areas for observation of students in real time or by videotape needs to be adjacent or in close proximity.

The EVMS Theresa A. Thomas Professional Skills Teaching and Assessment Center currently serves as a laboratory for simulation-based training. A recent study conducted at the Center looked at the external validation of resident physician evaluation using standardized patients (SPs). This "stealth patient" project, compared the testing of resident physician clinical abilities using two modalities with SPs. The study compared the performance of four PGY 3 and 4 resident physicians in the dermatology training program depending on whether they encountered a SP first in a Clinical Skills Assessment examination at the Thomas Center or in a standard medical office practice.

(k) Develop a conceptual model of high performance skill medical training simulators based on the training simulator requirements.

Research has proceeded so well that the development of high performance skill medical training simulators has resulted in the construction of working prototypes of both the wound debridement simulator and the augmented standardized patient simulator.

(l) Develop a research plan that will support the high performance skill medical training simulator development, including both the technologies to support the medical training simulator and the technology and methodology to support effective training tasks and transference of those to the real-world.

The wound debridement simulator includes a training component. The wound debridement trainer comprises a multimedia overview; a surgical simulator (tutorial component); and an assessment component. The initial training module provides an introduction and overview of the debridement process including information on procedures, instruments and wound types. This multimedia (web-based) component can launch a tutorial in the (second) simulator component or can run independently of the simulator since it requires only a web browser.

The surgical simulator provides the opportunity to prepare the site by scrubbing and rinsing the wound, to remove foreign objects, and to perform debridement on a virtual patient. Trainees must also learn to avoid cutting critical structures including blood vessels and nerves. Our simulator aims to deliver highly realistic graphics, real-time performance feedback, and haptic response to the trainee.

For the augmented standardized patient project, the use case developed envisions the simulator to be utilized initially in medical education research for determining its ability to provide improved instruction in conducting a physical examination of a patient versus the currently utilized approach. A research project involving the Objective Structured Clinical Examination (OSCE) required of all fourth-year medical students utilizing the augmented standardized patient technology is currently being designed for implementation in the summer of 2005. The outcomes of this study will drive further use case development in the areas of planning and training for a variety of health professions disciplines.

(m) Identify and explore haptics, auditory, simulation and visualization technologies that are available to implement high performance skill medical training simulators, including identification of a core set of technologies that are readily available and could be used to support a medical training simulator in the first quarter of calendar year 2005.

The literature on high performance skill acquisition was reviewed by human factors specialists prior to the development of the wound debridement and augmented standardized patient high performance skill medical simulators. In support of the wound debridement simulation development effort, an extensive literature search has been conducted in various areas of tissue modeling including the review of over 150 scholarly works. Simulation requirements currently under research and development include the following:

- Tissue deformation modeling using both mass-spring models and accelerated finite element modeling;
- Collision detection and cutting;
- Soft-body haptics rendering for two devices in the same space (one for each hand); and
- Realistic graphics rendering and visualization of anatomical data from the National Library of Medicine's Visible Human project to ensure correct anatomies and wounds.

For the augmented standardized patient simulator, a significant amount of research was conducted into sensor technology and capabilities, into three-dimensional positional tracking devices based on both magnetic and infrared technologies, into electronic stethoscope technology, into the processing of sound files and into those elements of the physical examination that needed to be replicated with the augmented standardized patient.

(n) Develop detailed high performance skill medical training simulators system design based on the conceptual model that will meet the identified simulator requirements.

This has been accomplished and construction of working prototypes of the two high performance skill medical training simulators has been completed.

(o) Construct high performance skill medical training simulators that meet the detailed design requirements and are supported by the research plan taking into account the core set of technologies previously identified.

Research has proceeded so well that the development of high performance skill medical training simulators has resulted in the construction of working prototypes of both the wound debridement simulator and the augmented standardized patient simulator.

(p) Validate the medical training simulators based on human factors principles of skill acquisition utilizing medical, physician's assistant, and psychology students.

The high performance skill medical simulators being developed under this research program are the wound debridement simulator and the augmented standardized patient simulator. These simulators are both currently in prototype stage. Human factors principles of skill acquisition have been included in the conceptual model of the simulators during the development phase. Validation experiments will commence after the designs of the high performance skill medical simulators are finalized and a final prototype is developed.

TASK 4

REGIONAL MEDICAL RESPONSE SIMULATIONS

FINAL TECHNICAL REPORT

Task 4 - Regional Medical Response Simulations.

The contractor shall:

(a) Identify requirements for an early warning system on changes in health status that could be used in a simulation of the Hampton Roads Metropolitan Medical Response System (MMRS).

Requirements for an early warning system were identified and used to support other requirements in this task. Specifically, the requirements for an early warning system of a population model underlying and supporting mass casualty simulation and analysis were identified.

In addition, EVMS has partnered its expertise in epidemiology and virology with the Virginia Modeling, Analysis and Simulation Center's expertise in large-scale population simulation and submitted a proposal to the National Science Foundation for a competitive multi-scale modeling grant. If funded, the grant would allow EVMS to study pathogen transmission and evolution in a simulated host population of at least 50,000 autonomous "intelligent-agents" with dynamic immune response. The simulations would be run on VMASC's large Beowulf cluster. The proposed pathogen model for the project is an enteric virus with a single-stranded RNA genome that would have a high rate of mutation and a great deal of antigenic diversity.

(b) Solicit and identify system use cases that will describe the intended use of the system to support planning, training, and decision support for mass casualty events.

Simulation system use cases have been identified and system level requirements have been developed. Specific population modeling requirements were devised to ensure that population behavior was captured at the appropriate level to achieve the overall goals of meeting the training and analysis requirements of health system managers.

(c) Develop detailed system-level requirements based on the identified use cases.

System level requirements have been identified. Please refer to the research paper that has been attached in Appendix D.1.

(d) Develop a conceptual model of the simulation system based on the system-level requirements.

The conceptual model design was completed. The model design requirements included the need to model the behavior of the population at the individual person level, to accurately portray regional population demographics, to portray the daily and weekly routines of individuals, to portray the social structure of individuals, to use an object-oriented design, to provide for model scalability and to model the effects of various types of mass casualty events on the health and behavior of individuals in a population. The derivation of a conceptual model based on the system-level requirements is in progress along with a supporting research and development plan. Please refer to the research paper that has been attached in Appendix D.1.

(e) Develop a research plan that will support simulation system and medical response system development.

Technical issues on the feasibility of implementing the detailed system design were researched. The supporting research and development plan is complete. Please refer to the research paper that has been attached in Appendix D.1.

(f) Develop the technologies required to support the simulation system, as well as the technology and methodology to support effective planning, training and decision support tasks.

Technologies required to support the simulation system have been identified. The technologies to support effective planning, training and decision support tasks will be acquired from third parties.

(g) Identify and explore simulation and visualization technologies that are available to implement the conceptual system mode, including identification of a core set of technologies that are readily available and could be used to support a medical response training exercise in calendar year 2004.

The applicable simulation and visualization technologies have been identified to support development of the system.

(h) Develop a detailed simulation system design based on the conceptual model that will meet the identified system-level requirements.

Research was conducted on the computing architecture necessary to implement the model. The software for the computing architecture was installed and made operational. The overall model control structure was designed and model behavior outlines were developed.

(i) Construct a simulation system that meets the detailed design requirements and is supported by the research plan taking into account the core set of technologies previously identified.

The base structures for holding model data were coded. In addition, the public health subject matter expert's comments have been incorporated in the system design. The simulation system itself remains under development.

(j) Validate the simulation system using accepted simulation system validation techniques.

A model-testing interface to monitor model output was implemented. Simulation validation will occur upon implementation of the simulation system.

FINDINGS/RESULTS

FINDINGS/RESULTS

Simulator Validation

The simulator validation studies conducted by the NCCMMS under this project determined that some medical simulations, albeit low-tech in nature, may be more effective than other similar high-tech simulations. This was demonstrated in the intravenous cannulation study in which the virtual reality simulator, while effective, was less effective in training this procedural skill than the more traditional type of training employing a simulated limb.

Technology Development

New medical and surgical simulators under development by the NCCMMS may be useful in medical and surgical education. For example, the augmented standardized patient simulator that is currently under development by the NCCMMS has the potential to revolutionize medical education by greatly expanding the range of pathologies that can be portrayed by the standardized patients utilized in the vast majority of U.S. and Canadian medical schools. The wound debridement simulator under development by the NCCMMS will allow for the rapid, just-in-time training of those physicians, often lacking recent or any wound debridement skills, to refresh or acquire these skills just prior to deployment to battlefield medical facilities.

Simulators and Curriculum Integration

The NCCMMS has determined that simulation has an appropriate place in medical education curricula by identifying, in typical undergraduate and graduate level medical education curricula, as well as in allied health curricula, appropriate uses for medical modeling and simulation resources. In addition, the NCCMMS has determined that up to the present there has been no single repository of medical modeling and simulation data to advance the field. The Medical Modeling and Simulation Database developed by the NCCMMS has addressed this need.

Regional Medical Response Simulations

The NCMMS has determined that it is feasible to develop a simulation of the effects of a mass casualty event on the medical infrastructure on a geographical region, in this particular case the multi-jurisdictional region of eastern Virginia. This determination will ultimately result in a simulation that will be utilizable and customizable by other regions to simulate the effects of such an incident on their own regions.

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

Additional multi-center research on the validation of medical models and simulations is necessary to conclusively validate medical models and simulations. Extensive time restraints on the utilization of medical students and medical residents in medical modeling and simulation validation research resulted in a smaller number of research subjects in the studies conducted by the NCCMMS than what was originally envisioned and intended.

While their potential for contributing to medical and surgical education is significant, additional research will be necessary to further develop the wound debridement and augmented standardized patient simulators beyond the somewhat crude prototypes that have been developed to this point.

Incorporating medical models and simulations in medical education curricula is difficult in a time- and resource-restrained environment with multiple administrative entities involved. Without intra- and extra-departmental champions for incorporating modeling and simulation in the respective departments' curricula, such changes are difficult to implement.

It has been possible to develop a comprehensive medical modeling and simulation database to further the field. It is a challenge to keep such a database current given the recent tremendous growth of the field. It may be necessary for the NCCMMS to partner with other interested organizations to provide sufficient resources to sustain and improve the database in the future.

Additional research is necessary to more comprehensively model the medical effects of a mass casualty incident in a multi-jurisdictional region such as Hampton Roads, Virginia.

RECOMMENDATIONS

Future medical modeling and simulation validation research studies need to incorporate additional subjects among multiple research centers to obtain sufficient numbers of research subjects. While this will most likely entail additional costs, it must be acknowledged that the cost of acquiring medical models and simulators for the military services and for medical education institutions such as medical schools is not insignificant and it is key that models and simulators on the market be exhaustively studied and validated for effectiveness prior to their purchase by these types of entities.

More realistic timetables for the development of new medical models and simulations should be adopted by researchers that reflect the complexity of this task. It has proven very difficult to find dedicated time to bring the panoply of medical and engineering researchers together for developing complex simulators such as the augmented standardized patient and the wound debridement simulators. While considerable progress has been made in developing the prototypes of these two simulators, not as much progress was achieved during the project period as was initially anticipated.

Medical education curricular redesign should incorporate validated medical modeling and simulation tools. Increasing numbers of medical models and simulations are being validated as effective. These tools should be incorporated in medical education curricula to improve teaching effectiveness.

The Medical Modeling and Simulation Database developed by the NCCMMS should be updated on a regular basis to maintain its usefulness for researchers and others in the field. This may require a strategic partnership of the NCCMMS with other entities to provide the necessary resources.

Additional research on the simulation of medical effects associated with mass casualty incidents needs to be undertaken as these types of simulations are very complex and as new threats develop and available medical resources evolve.

Appendix A: Publications Supporting Task 1 – Simulator Validation

Several publications were either presented or prepared during the period of performance. These have been referred in the report above and are summarized in the following subsections. The five publications that appear in the following subsections are:

1. Mark W. Scerbo, James P. Bliss, Elizabeth A. Schmidt, Sommer N. Thompson, Theresa D. Cox, and Harry J. Poland, "A Comparison of the CathSim™ System and Simulated Limbs for Teaching Intravenous Cannulation," In J.D. Westwood et al. (Eds.), *Medicine Meets Virtual Reality*, 12, (196-198). Amsterdam: IOS Press, 2004.
2. Mark W. Scerbo, James P. Bliss, Elizabeth A. Schmidt, & Sommer N. Thompson, "The Efficacy of a Medical Virtual Reality Simulator for Training Phlebotomy," to appear in *Human Factors.*
3. Elizabeth A. Schmidt, Mark W. Scerbo, James P. Bliss, and Sommer N. Thompson, "Skill Acquisition with a VR Simulator for Phlebotomy," *Second Conference on Human Performance, Situation Awareness, and Automation*, Daytona Beach, FL., Mar., 2004.
4. James P. Bliss, Hope S. Hanner-Bailey, & Mark W. Scerbo, "Determining the Efficacy of an Immersive Trainer for Arthroscopy Skills," presented at *Medicine Meets Virtual Reality XIII*, Long Beach, CA, Jan., 2005.
5. Mark W. Bowyer, Elisabeth A. Pimentel, Jennifer B. Fellows, Ryan L. Scofield, Vincent L. Ackerman, Patrick E. Horne, Alan V. Liu, Gerald R. Schwartz, and Mark W. Scerbo, "Teaching Intravenous Cannulation to Medical Students: Comparative Analysis of Two Simulators and Two Traditional Educational Approaches," In J.D. Westwood et. al. (Eds.), *Medicine Meets Virtual Reality*, 13, (57-63). Amsterdam: IOS Press, 2005.

A.1 Mark W. Scerbo, James P. Bliss, Elizabeth A. Schmidt, Sommer N. Thompson, Theresa D. Cox, and Harry J. Poland, "A Comparison of the CathSim™ System and Simulated Limbs for Teaching Intravenous Cannulation," In J.D. Westwood et al. (Eds.), *Medicine Meets Virtual Reality*, 12, (196-198). Amsterdam: IOS Press, 2004.

In J. D. Westwood et al. (Eds.), *Medicine meets virtual reality, 12*, (196-198). Amsterdam: IOS Press, 2004.

A Comparison of the CathSim™ System and Simulated Limbs for Teaching Intravenous Cannulation

Mark W. SCERBO, James P. BLISS, Elizabeth A. SCHMIDT, Sommer N. THOMPSON
Old Dominion University
Theresa D. COX, and Harry J. POLAND
Eastern Virginia Medical School

Abstract. The present study describes a comparison between the CathSim™ VR simulator and simulated limbs for training IV cannulation. Two groups of physician assistant students underwent 2 hours of training on either method. Performance was assessed before and after training with a standardized assessment form. The results showed that all students improved after training, but the degree of improvement was greater for those trained with the simulated limbs. These findings may be due to differences between the two training methods as well as the methodology adopted in the present study.

1. Background

The use of high fidelity virtual reality simulators has had a profound impact on operator performance in aerospace and the military for decades. In medicine, the first virtual reality simulators began to appear in the early 1990s [1]. Since then, there has been a proliferation of systems to address a variety of procedures, but there have been few attempts to validate their effectiveness [2, 3]. The present study was designed to compare the effectiveness of a VR system for teaching intravenous (IV) cannulation with a more traditional method using simulated limbs.

1.2 Research on VR systems for IV Procedures

Few studies have been published that examine the effectiveness of VR systems for teaching IV procedures and in those studies, the demonstrated benefits of VR training have been limited. For instance, Prystowsky et al. [4] studied IV catheterization with a VR system incorporating stereoscopic views and haptic feedback. They studied first- and third-year medical students and surgical residents. Their participants viewed a 5-min training video and then made 2 attempts at IV insertion on one another. Afterward, they practiced with the simulator for 12 minutes. They then made 2 more attempts at cannulation on another participant. Prystowsky et al. found no differences in performance between pretest and posttest scores for any group of participants. Success rates were tied only to experience. Participants with more formal training (and experience with the cannulation procedure) performed better than less experienced participants.

In another study, Chang, Chung, and Wong [5] trained a group of nurses using either the CathSim™ system (see description below) or a plastic arm. One group was trained for 2 hours with the plastic arm. The other group was given a 2-hour introductory lesson on the CathSim™ system and then left to practice on their own during the following week. All participants were then assessed on their first attempt at cannulation with a genuine patient. The results showed that most participants were successful on their first attempt, but the success rate was slightly higher for those trained on the plastic arm. There were no differences between the two groups on the assessment scores. Chang and her colleagues argued that the higher success rates observed with participants in the plastic arm group may have been due to those participants having more experience with phlebotomy. They also indicated that participants in the CathSim™ group had not spent sufficient time practicing. Unfortunately, Chang et al. did not standardize the practice time for all CathSim™ participants, nor did they report the amount of time the CathSim participants spent practicing on their own.

1.3 The Present Study

The present study is the first in a series to examine the effectiveness of a VR simulator for IV procedures. The results of earlier studies failed to show any benefits of VR training. At present, it is not known whether the lack of success is due to the nature of the VR simulator itself or to the methods used to evaluate the simulator. For example, Prystowsky et al. [4] did not compare instructional methods. Thus, one cannot say how training on their simulator might compare to other techniques. Chang et al. [5] also observed no benefits of VR training, but they did not standardize the training for individuals in the VR condition.

The goal of the present study, therefore, was to reexamine training methods for IV cannulation. Training on a VR system was compared to a more traditional method using a simulated limb. Further, training was standardized across conditions and participation was limited to individuals who had no prior experience with the procedure. In addition, all participants were assessed before and after training using a standardized instrument. Also, because the earlier studies produced null results another goal of the present study was to examine performance at a more detailed level by addressing critical steps of the procedure common to both methods.

2. Methodology

2.1 Participants

Participants were 16 first-year students in the Master of Physician Assistant program at Eastern Virginia Medical School. The students participated as part of their course requirements. They ranged in age from 21 to 38 (M= 26).

All participants had some previous medical experience (e.g., working in a nursing home, as an EMT volunteer, or as a pediatric assistant), but few had any exposure to IV procedures. Although two participants indicated they had been exposed to IV procedures at a retreat/workshop, none of the participants had any experience performing IV cannulation. Only one participant reported experience with other types of subcutaneous or intramuscular injections. No participants reported experience with any other type of medical simulator, but all were regular computer users.

2.2 Materials

2.2.1 Simulated Arms

A Life/Form® simulated arm was used for the pretest and posttest. The simulated arm has a layer of vinyl skin covering a network of latex veins. The simulated arm has three main puncture areas for needle insertion: on the back of the hand, on the antecubital fossa, and along the wrist and forearm. A container of artificial blood is suspended over the arm and connected to the arm with IV tubing. Gravity draws the artificial blood into the system of veins and provides flashback when the vein is punctured with a needle. Other materials used in conjunction with the simulated arm were standard surgical gloves, tourniquets, alcohol swabs, iodine swabs, gauze pads, 20-gauge "over the needle" IV catheter needles, and short IV extension catheter tubing to create a saline loc. Upon completion of the procedure, all needles were disposed of in a biohazard sharps container.

2.2.2 The VR System

The VR simulator used in the present study was the CathSim™ system available from Immersion Medical, Inc. The CathSim™ system is a microcomputer-based simulator originally developed by HT Medical Systems, Inc., in collaboration with the Department of Nursing, Food and Nutrition of Plattsburgh State University of New York [6, 7]. The CathSim™ system provides training on a subset of needle stick procedures including IV catheterization and phlebotomy. The physical system consists of an IBM-compatible computer accompanied by an AccuTouch™ six-degree-of-freedom haptic feedback device that simulates the catheter needle/hub assembly and a section of the skin for traction. The AccuTouch™ device is designed to allow trainees to experience the feeling associated with inserting a needle into the skin and vein.

The system includes six case scenarios for teaching catheterization and phlebotomy. The case patients differ in difficulty. For instance, there is an adult male with no complications and pediatric and geriatric cases with varying complications. The system provides tutorial video clips to familiarize students about many procedural details; however, some background knowledge of procedures is required (e.g., how to select an appropriate needle gauge and knowledge of potential contraindications).

Students select a patient with the computer mouse. After choosing a case, they are shown an image of the patient's arm on the computer screen and must then select a site for insertion. Next, the student uses the computer's mouse to apply a tourniquet, palpate the vein, and cleanse the site by pointing, clicking, and dragging objects on the screen. After the site has been prepared, the student selects the appropriate gauge needle and then manipulates the mouse and the AccuTouch™ device to position the needle and apply skin traction. Skin traction is simulated by pulling down on the rubber-stripping portion of the device with the thumb. Next, the needle is inserted by first fully retracting the arm assembly of the Accutouch™ device and then pushing forward to insert it back into the device while simultaneously looking at the computer monitor to ensure vein access and confirmation of blood back flow. The student then completes the procedure and withdraws the needle. The system automatically records many performance metrics, however, only four metrics were analyzed in the present study: 1) procedure success, 2) pain factor (1-10, with lower numbers reflecting less pain), 3) angle of needle insertion, and 4) tourniquet release time.

2.3 Procedure

All participants received a lecture as part of their normal curriculum on IV cannulation. Following the lecture, all participants were asked to complete a background questionnaire. Upon completion of the questionnaire, half of the participants were assigned at random to the traditional training group and the other half to the VR training group.

All participants then completed a cannulation pretest on the simulated arm. Their performance was assessed with a modified version of the standard instrument used to certify students on this procedure. The instrument was modified to include several rating scales so that performance on some steps could be assessed in qualitative terms.

The next two weeks consisted of training. All participants were scheduled for two 2-hour labs held a week apart. Participants in the traditional training group met in their classroom during their scheduled labs and practiced the procedure with the simulated arms under the supervision of the course instructor.

The VR participants worked in pairs and practiced with the CathSim™ system. The participants practiced individually and took alternating 5-10 minute turns. In total, each participant received about 50 minutes of practice during each lab. The VR participants were not permitted to use any of the additional training aids offered by CathSim™ (i.e., the side or transparency views). Participants worked with only the practice male, teenage female, young adult male, and obese female case patients, but performed the IV insertions at all three possible sites. The VR participants were required to perform a successful procedure with no hematoma and achieve a pain factor of 3 or less and a tourniquet time of less than two minutes in order to move on to another simulated patient.

Upon completion of training, all participants performed a posttest on the simulated arms. Their performance was assessed with the same instrument used for the pretest. Afterward, participants in both groups were offered the opportunity to train with the other method.

3. Results and Discussion

One student did not complete the posttest, so data from only fifteen participants were analyzed. The results of the pretest and posttest are shown in the Full Procedure section of Table 1. Both groups performed similarly on the pretest. Scores on the pretest and posttest showed a significant improvement for the traditional group $t\ (6) = 4.4$, $p < .005$ as well as the VR system group $t\ (7) = 3.08$, $p < .01$. Further, the posttest scores for the traditional group were significantly higher than those for the VR system group $t\ (13) = 3.08$, $p < .01$ (using a two-tailed test).

Table 1. Mean Assessment Scores on the Pretest and Posttest for Both Groups on the Full Procedure and on Selected Steps

	Full Procedure		Common Steps	
	Pretest	Posttest	Pretest	Posttest
Traditional	34.75 (15.89)	55.43 (6.6)	23.71 (9.41)	42.00 (2.62)
VR System	29.5 (9.12)	43.25 (8.41)	24.00 (5.00)	34.86 (7.47)

(Standard deviations in parentheses)

For those individuals working with the VR system, an analysis of training performance showed improvements on the four critical metrics. Specifically, participants completed more procedures, achieved less hematoma, produced less pain, and completed the procedure more quickly during their second session (see Table 2).

Table 2. Performance Metrics for the VR System Participants on Each Training Session

Metric	Session 1	Session 2
Percent of Successful Procedures	58.8	82.8
Mean Procedure Time (in sec.)	110.2 (65.03)	77.99 (34.86)
Percent Hematoma	31.4	14.1
Mean Pain Score	5.98 (3.82)	3.76 (3.15)

(Standard deviations in parentheses)

The results of the present study indicate that training was effective. Students improved under both methods. However, the degree of improvement was greater with the traditional method using simulated limbs. In this regard, these findings are similar to those of others who have also observed limited success with VR systems for IV cannulation [4, 5].

The results of this and other studies may be due to several important and unique differences between the two training methods. It is important to understand that both devices are designed to provide training on the IV cannulation procedure. However, each device affords the opportunity to practice only a subset of the steps required to successfully perform the entire procedure. Further, each device differs in the number of steps it addresses (see Table 3). Moreover, the opportunity to perform those steps with either device may differ qualitatively from performing those steps with a genuine patient.

The steps required to perform an IV cannulation are listed in Table 3. Of the 25 steps listed, 20 can be practiced with the simulated arm but only 13 can be practiced with the CathSim™ system. For example, both devices allow the user to practice inserting a needle (Step 16), however, only the simulated arm allows the user to practice connecting the IV tubing to the catheter (Step 20). Neither method allows the student to reexamine the cannulation site for postprocedure complications (Step 25). Further, those steps marked with an asterisk cannot be performed as they would with a genuine patient. Thus, for instance, neither system provides complete freedom in selecting an insertion site. Similarly, although the CathSim™ system requires the user to apply a tourniquet, the step is accomplished using the computer's mouse. The user cannot practice tying the tourniquet around a patient's arm.

Given these differences, one could argue that the higher scores observed with the simulated arms occurred because those students had the opportunity to practice more steps of the full procedure. To address this possibility, a second analysis was performed on only the 12 steps common to both methods (those steps marked with both an S and C in Table 3). Those data appear under the Common Steps section of Table 2. As can be seen in the table, the pattern is similar to that of the full procedure. Both groups performed comparably on the common steps during the pretest. Both groups improved after training, but again the greatest improvement occurred for the students who trained with the simulated limb.

Table 3. Steps of the IV Cannulation Procedure that can be Practiced using the Simulated Arm (S) and CathSim System (C)

		1. Introduce yourself to patient
		2. Verify patient's identification
S		3. Gather and arrange all materials (tourniquet, catheter, gauze pads, etc.)
S		4. Set aside four 6 in tape strips to affix IV tubing later on
S		5. Draw saline/distilled water into syringe and replace cap
		6. Wash hands
S	C*	7. Put on gloves
S*	C*	8. Inspect area for cannulation site in vein
S	C*	9. Put on tourniquet (3-4 inches above insertion site)
	C*	10. Palpate vein
S	C	11. Swab area with alcohol pad in concentric circles moving outward, let air dry
S	C	12. Grasp cannula by flashback chamber behind needle hub
S		13. Remove protective sleeve
S	C*	14. Stabilize vein
S	C*	15. Hold needle at 15-30 degree angle, bevel side up
S	C	16. Insert needle
S	C	17. Advance needle into vein
S	C	18. Check for back flow of blood
S	C	19. Advance catheter and withdraw needle
S		20. Connect IV tubing to catheter
S	C	21. Release tourniquet
S		22. Push saline into vein
S		23. Affix catheter and tubing to arm
S		24. Label dressing
		25. Recheck cannulation site for postprocedure complications

* indicates limited or partial ability to practice the step

4. Conclusion

The pattern of results obtained in the present study suggest that using the CathSim™ VR system to teach IV cannulation is less effective than using simulated limbs. One explanation for these differences may lie with the unique characteristics of each training method. As noted above, each method addresses a different set of steps from the full procedure. Moreover, the activities required to perform several steps on the CathSim™ system do not faithfully reproduce the activities necessary for implementing those steps on a simulated arm or on a genuine patient.

There are, however, several methodological aspects of this study that may have also contributed to these results. First, the students who trained with the simulated limbs may have had an advantage over the other group because the pretest and posttest were conducted using same system on which they practiced. Therefore, it is possible that the skills acquired by students in the CathSim™ group did not transfer fully to the simulated arm. Although this is a possibility, one would hope that skills acquired with the CathSim™ system would generalize beyond the simulator and indeed transfer to genuine patients. A second issue concerns the training times. In the present study, the session times were standardized across conditions. Although this control was necessary to compare

instructional methods, it does not indicate how much additional training on the CathSim™ system would be needed to achieve similar levels of performance. Consequently, a second study is currently underway to address these issues.

Acknowledgements

This work was supported by the National Center for Collaboration in Medical Modeling and Simulation, a joint enterprise of the Eastern Virginia Medical School and the Virginia Modeling, Analysis and Simulation Center (VMASC) at Old Dominion University. Partial funding for this study was provided by the Telemedicine and Advanced Technologies Research Center, U.S. Army Medical Research and Material Command, contract DAMD17-03-2-0004. The ideas and opinions presented in this paper represent the views of the authors and do not necessarily represent the views of the Department of Defense.

References

[1] Satava, R. M. (1993). Virtual reality surgical simulator: The first steps. *Surgical Endoscopy, 7*, 203-205.

[2] Magee, J. H. (2003). Validation of medical modeling & simulation training devices and systems. In J. D. Westwood et al. (Eds.), *Medicine meets virtual reality, 11*, (196-198). Amsterdam: IOS Press.

[3] Satava, R. M. (2001). Accomplishments and challenges of surgical simulation: Dawning of the next-generation surgical education. *Surgical Endoscopy, 15*, 232-241.

[4] Prystowsky, J.B., Regehr, G., Rogers, D.A., Loan, J.P., Hiemenz, L.L., & Smith, K.M. (1999). A virtual reality module for intravenous catheter placement. *The American Journal of Surgery, 177,* 171-175

[5] Chang, K. K., Chung, J. W., & Wong, T. K. (2002). Learning intravenous cannulation: A comparison of the conventional method and the CathSim intravenous training system. *Journal of Clinical Nursing, 11*, 73-78.

[6] Barker, V. L. (1999). CathSim™. In J. D. Westwood et al. (Eds.), *Medicine meets virtual reality,* (36-37). Amsterdam: IOS Press.

[7] Ursino, M., Tasto, J. L., Nguyen, B.H., Cunningham, R., & Merril, G. L. (1999). CathSim™: An intravenous catheterization simulator on a PC. In J. D. Westwood et al. (Eds.), *Medicine meets virtual reality,* (360-366). Amsterdam: IOS Press.

A.2 Mark W. Scerbo, James P. Bliss, Elizabeth A. Schmidt, & Sommer N. Thompson, "The Efficacy of a Medical Virtual Reality Simulator for Training Phlebotomy," to appear in *Human Factors.*

To appear in *Human Factors*

The Efficacy of a Medical Virtual Reality Simulator for Training Phlebotomy

Mark W. Scerbo, James P. Bliss, Elizabeth A. Schmidt, & Sommer N. Thompson

Old Dominion University

Norfolk, VA

Keywords: Virtual reality, medicine, phlebotomy, simulation, training

Abstract

Phlebotomy, or drawing blood, is one of the most common medical procedures; yet, there are no universal standards for training and assessing performance. This can ultimately lead to injuries and inaccurate test results if the procedure is improperly performed. The present study compared the effectiveness of a virtual reality (VR) simulator for training phlebotomy with a more traditional approach using simulated limbs. Twenty third-year medical students were trained under one of the two methods and had their performance assessed with a 28-item checklist. The results showed that performance improvements were limited to those who trained with the simulated limbs and a detailed comparison of the two systems revealed several functional and physical differences that may explain these findings. Although participants trained with simulated limbs performed better, the metrics recorded by the VR system may address other aspects of performance that could be beneficial. The present study highlights the potential for medical simulators to improve patient safety by enabling trainees to practice procedures on devices instead of patients. Potential applications of this research include training and performance assessment.

The Efficacy of a Medical Virtual Reality Simulator for Training Phlebotomy

Simulators have been a standard component of training for a variety of jobs including military strategic and tactical command operations, managerial decision-making, nuclear power plant operations, and space flight operations. The area in which simulators have had perhaps the most significant impact on training is aviation, where flight simulators have been in use for over 60 years. The current class of high fidelity flight training devices can reproduce almost all of the essential characteristics of piloting an aircraft including cockpit displays and controls, out-the-window scenery, radio communications, various weather and traffic conditions, and in some cases even the appropriate pitch, roll, and yaw movements of the aircraft. In fact, it is not uncommon for flight schools to require a minimum number of hours with a simulator to obtain a pilot's license. One could easily argue that the availability of simulators for training has played a critical role in the evolution of military and commercial aviation.

Medical Virtual Reality Simulators

Although the necessity of simulators for training pilots seems obvious, simulators have not played a significant role in the training and education of medical professionals. Medical education has traditionally followed an apprenticeship model where procedures are learned by the "see one, do one, teach one" approach. In fact, Dawson (2002) notes that this approach to medical education has remained constant since ancient Egyptian times.

Medical simulators are relatively new. Although medical simulation *devices* have been around since the 1940s, most of them have been physical models with very limited functionality. Since their inception, medical simulators have evolved along two paths. One class of simulators is mannequin-based and provides a physical model of the patient as well as physiological responses to treatments (Dawson, 2002). The second class features virtual reality (VR) technology. These

simulators incorporate 3-dimensional graphics of patients or internal organs and provide computer-based training on a variety of procedures including cholecystectomy (i.e., gall bladder removal), arthroscopy (e.g., knee surgery), upper and lower endoscopy (e.g., colonoscopy), and vascular access such as phlebotomy (i.e., drawing blood) and intravenous (IV) catheter insertion (Satava, 2001). It is this second class of simulators that is the focus of the present paper.

The first VR-based surgical simulator appeared in the early 1990s (Satava, 1993) and provided a 3-dimensional representation of internal organs in the upper abdomen that could be observed from different perspectives. Shortly thereafter, systems emerged that incorporated accurate representations of organs based on computed tomography (CT) scans. By the end of the 1990s, systems were being developed that addressed specific procedures including knee surgery, ophthalmology, and sinus surgery (Satava, 2001).

One of the important factors in the rapid evolution of this technology, and one that made medical VR systems more commercially viable, was the addition of haptic interfaces. Medical professionals often rely on the sense of touch for performing procedures. For example, nurses use their fingers to take a patient's pulse. Surgeons use instruments to probe, cut, and suture organs. Haptic systems including force feedback allow users to touch virtual objects. The user holds facsimiles of medical instruments in their hands. The handles, grips, and controls are identical to the actual tools, but the operational ends of the tools are encased in a housing and modified with a force feedback system that monitors hand movements in 3-dimensional space to detect collisions with virtual objects and surfaces (e.g., skin, organs, etc.) and to provide appropriate levels of force to impede further movement (see Basdogan & Srinivasan, 2002). Consequently, it feels like the instruments are contacting, probing, and penetrating organic structures.

There are several ways that medical simulators can help to improve patient safety. First, and most important, they allow medical personnel to train on devices instead of practicing on actual patients. This allows trainees to learn from their mistakes without injuring anyone and to meet minimal levels of acceptable performance before seeing a patient. Second, simulators also

offer an environment to train specific skills in the absence of uncontrollable influences (e.g., variations in comorbidity of condition). Aside from these obvious advantages to patient safety, medical simulators offer other benefits. For instance, they can provide trainees with an unlimited number of trials to acquire skills and give immediate feedback on their performance. They can also expose medical students to rare or infrequent conditions. Last, they can help decrease the dependence on animals and cadavers to practice procedures.

A VR System for IV Procedures

A VR simulator was recently developed for IV procedures (see Ursino, Tasto, Nguyen, Cunningham, & Merril, 1999) and is available from Immersion Medical, Inc. The CathSim™ system provides training on a subset of needle stick procedures including catheter insertion (cannulation) and phlebotomy. Although both procedures require a needle to gain access to the underlying vascular system, the objective of each procedure is quite different. In cannulation, a needle is used to place a catheter into the patient's vein so that fluids may be introduced to the body through that portal. By contrast, the purpose of phlebotomy is to take blood samples from the body. Thus, a needle assembly is inserted into a vein and blood is withdrawn into one or more collection tubes.

CathSim™ is a multimedia system that simulates both IV procedures with visual, auditory, and haptic displays. The physical system consists of an IBM-compatible computer accompanied by an AccuTouch™ six-degree-of-freedom haptic feedback device that simulates the needle and a vinyl strip that simulates the skin. The AccuTouch™ device allows trainees to experience force feedback associated with inserting a needle into the skin and vein.

The system includes a variety of unique case patient scenarios. For example, there is an adult male with no complications and pediatric and geriatric cases with varying complications. The system provides tutorial video clips to familiarize students about many procedural details;

however, some background knowledge of procedures is required (e.g., how to select an appropriate needle gauge and knowledge of potential contraindications).

The user operates the system by making selections with the computer's mouse and manipulating the AccuTouch™ device to insert the needle. The system records many performance metrics that can provide trainees with immediate feedback to improve and refine their performance. These metrics include procedure success, angle of needle insertion, occurrence of hematoma, session time, tourniquet placement time, and a pain factor. The CathSim™ system also incorporates two visual training aids that allow students to see a representation of the needle entering the arm from a side view or a "transparent" view of the underlying vascular system.

To date, there have been few published studies examining the effectiveness of VR systems for teaching IV procedures. Prystowsky, et al. (1999) studied first- and third-year medical students and surgical residents with a VR system for cannulation that incorporated stereoscopic views and haptic feedback. They gave their participants background information about the procedure and then required them to attempt a cannulation on each other. The participants then practiced with the simulator for 12 minutes and subsequently attempted a second cannulation on each other. Performance ratings revealed no differences between the first and second attempt for any group of participants. Reznek, Rawn, and Krummel (2002) conducted a similar study using the CathSim™ VR system, but found differences related to only years of medical training. Chang, Chung, and Wong (2002) compared the CathSim™ system and plastic arms for teaching cannulation to student nurses. The participants were given two hours of training with either method and were assessed on their first attempt at cannulation with a genuine patient. Chang and her colleagues observed that most participants were successful, but the success rate was slightly higher for those trained on the plastic arm. Interestingly, they suggested that the superior performance of participants in the plastic arm group might have been due to their greater experience with phlebotomy. More recently, Scerbo et al. (2004) conducted a similar study

comparing the CathSim™ system and simulated limbs for training IV cannulation. They gave two groups of physician assistant students two hours of training on either method. Performance was measured before and after training with a standardized assessment form. The results showed that all students improved after training, but that the degree of improvement was significantly greater for those trained with the simulated limbs. Engum, Jeffries, and Fisher (2003) also reported some advantages of the simulated limbs over the CathSim™ system.

The Present Study

The results of the studies described above suggest that VR training for IV catheterization procedures is not as effective as other methods; however, existing studies have addressed only IV cannulation. None has examined phlebotomy. This is an unfortunate oversight because phlebotomy is performed more often than cannulation by medical personnel in hospitals, physician's offices, clinics, and even in mobile blood vehicles. Further, there are a number of concerns with respect to patient safety. According to Mishori (2004), there are no licensure requirements for phlebotomists and the field is largely unregulated. Thus, the level of competency of those performing the procedure can vary widely. In fact, in one study 35% of patients surveyed reported discomfort with the procedure (Howanitz, Cembrowski, & Bachner, 1991). Further, injuries can and do occur. Patients have suffered punctured arteries, chronic pain from needle sticks through a nerve, and concussions and bone fractures from fainting spells that occurred when they were left unattended (Mishori, 2004). There are also indirect effects on patient safety. An improperly performed procedure can generate erroneous lab results that either leave existing conditions undetected or indicate the presence of non-existing conditions and subject the patient to additional unnecessary tests and associated risks (Howanitz, Cembrowski, & Bachner, 1991).

Aside from the patient safety issues, there are other factors that affect the ability to perform the procedure. Many issues can complicate the process. Patients vary widely in terms of their age, size, obesity, general health, and personal habits such as smoking or intravenous drug use. All of these factors affect vein accessibility and elasticity and can make the procedure more difficult to perform. Further, Clover (2002) argues that the safety risks for both the phlebotomist and the patient are often underrated. There are a variety of safety issues unique to phlebotomy concerning the practice of collecting multiple blood samples during a single procedure and the use of sharps disposable containers. Moreover, it is estimated that 600,000 to 800,000 needlestick injures occur in healthcare settings each year (NIOSH, 1999) and that a large percentage of these carry the ever-present threat of blood-borne diseases. Last, there are no performance-based standards for assessing phlebotomy skills in healthcare workers. All of these factors make phlebotomy a particularly good candidate for simulation-based training. Thus, the primary objective of the present study was to examine the effectiveness of a VR system for training phlebotomy.

At present, a standard method for teaching phlebotomy uses plastic simulated limbs; therefore, the goal of the present study was to compare VR-based CathSim™ training with the more traditional simulated limb training. It was expected that training under each method would improve performance; however, expectations regarding one method over another were unclear. On one hand, it is possible that the limited success observed in the previous studies with a VR system for cannulation would also be observed for phlebotomy. On the other hand, differences between the two procedures might make phlebotomy more amenable to VR simulation. In cannulation, once the needle and catheter are positioned, the catheter is slid off the needle into the vein and then the needle is withdrawn. By contrast, in the phlebotomy procedure once the needle is positioned, it must be held in place while the blood collection tubes are positioned, filled, and swapped. Thus, it is possible that needle placement and collection activities required for phlebotomy might be more easily achieved with the VR system.

The second objective was to examine the specific differences between the two training methods. Prystowsky et al. (1999) and Chang et al. (2002) did not offer detailed explanations for their findings. Scerbo et al. (2004), however, did describe differences between the two technologies they studied. The same procedure will be followed in this study. Toward that end, a physical and functional analysis of the two training systems will be performed to address any potential differences in training effectiveness.

Method

Participants.

Participants were 20 third-year medical students from Eastern Virginia Medical School in Norfolk, VA. The students participated as part of their family medicine clerkship (rotation) requirements. They ranged in age from 23 to 30 years (M= 25.45). All participants had some previous medical experience (e.g., working in a nursing home, serving as an EMT volunteer, working as a research assistant, hospital volunteer, or certified nurses aid). Four students indicated some exposure to phlebotomy procedures, but none had ever attempted one. Nine students indicated some experience with subcutaneous or intramuscular injections. No participant reported experience with any other type of medical VR simulator, but all used computers on a regular basis (approximately 13 hrs/wk on average for word processing, email, etc.) and three reported using other types of simulators (e.g., driving, games, etc.).

Materials

Simulated arm. A Life/Form® simulated arm was used for the pretest and posttest (see Figure 1). The simulated arm has a layer of vinyl skin covering a network of latex veins. A standard IV bag filled with a quart of artificial blood is hung 2 ft. above the arm and is connected to the tubing of the arm. Gravity draws the artificial blood into the system of veins and provides

flashback when the vein is punctured with a needle. The simulated arm has several insertion sites but only the antecubital fossa was used in this study. Other materials used in conjunction with the simulated arm were standard surgical gloves, tourniquets, alcohol swabs, gauze pads, 20-gauge needles, and various Vacutainer™ tubes and holders. Upon completion of the procedure, all needles were discarded in a biohazard sharps container.

CathSim™ system. The VR simulator used in the present study was the CathSim™ system from Immersion Medical, Inc. described earlier (see Figure 2). The phlebotomy module includes six case patient scenarios. Students select a patient with the computer mouse and are then shown an image of the patient's arm. Next, they use the mouse to select a site for needle insertion, to apply a tourniquet, palpate the vein, and cleanse the site by clicking and dragging objects on the screen. After the site has been prepared, the student manipulates the mouse and the AccuTouch™ device to position the needle. Skin traction is simulated by pulling down on the vinyl strip area of the device with the thumb. Next, the needle is inserted by fully retracting the arm assembly of the Accutouch™ device and then pushing it forward to insert it back into the device while viewing the computer monitor to ensure vein access. Once the needle is positioned properly, the mouse is used to select the correct order of collection tubes which are then inserted into and released from the Vacutainer™ holder component of the AccuTouch™ device. The student completes the procedure by withdrawing the needle and depositing it in a simulated biohazard sharps container.

Four system-generated performance metrics were recorded. The first was a pain factor with a range of 1-10 (lower numbers reflect less pain). The pain score increases as the needle penetrates the flesh to deeper levels and as the needle is rotated after it is embedded in the skin. The second metric was the presence of a hematoma that occurs if the needle passes entirely through the vein. Further injury increases the rate and intensity of the hematoma. The third metric was the duration that the tourniquet was in place and the fourth metric was the success/failure of

the procedure. None of the augmented view training aids (interior or side views of the underlying vascular system) was used.

Assessment

Performance was assessed using a checklist developed in the medical school to grade students on this procedure. The instrument addresses 28 specific steps of the procedure and provides scores for preparation, insertion, withdrawal/closure, and overall performance. The instrument was modified for the present study to include several rating scales so that performance on some steps could be assessed in a quantitative manner and weighted according to their importance. In addition, four steps were omitted because they could not be assessed with both simulators. The instrument had a range of 0 to 88 points with higher scores reflecting better performance.

Procedure

All participants were asked to complete a background questionnaire. Afterward, they were shown an instructional video on phlebotomy. Half of the participants were assigned at random to the simulated arm group and the other half to the VR group. However, due to some scheduling changes 9 students ended up in the VR group and 11 in the simulated arm group. Those participants who had indicated some exposure to phlebotomy were counterbalanced across the groups.

All participants then performed a phlebotomy pretest on the simulated arm. The next two weeks consisted of training. Participants were scheduled for two one-hour labs held one week apart. They were trained in individual sessions and received instruction from the experimenter as needed. Those in the simulated arm group practiced until they were able to perform a successful

procedure (as determined by the assessment form used for the pretest and posttest) with a tourniquet time of less than 90 seconds. Those in the VR group worked with the six different case patients. There were slight differences among the cases that made some more challenging than others (e.g., less visible veins or a steeper angle of needle insertion required to access the vein). Students began their training session with the easiest case and continued to practice with this case until they performed a successful procedure with no hematoma, a pain factor of 3 or less, and a tourniquet time of less than 90 seconds. Once they reached criterion they moved on to another case patient. The students practiced on the remaining case patients in order of increasing difficulty.

Upon completion of training, all participants performed a posttest on the simulated arm. Their performance was assessed with the same instrument used for the pretest.

Results

Full Procedure

Total scores on the assessment forms were computed for each student on the pretest and posttest. The means for each group are shown on the left side of Table 1. These scores were analyzed using a mixed-factor ANOVA with test as the within-subjects factor and group as the between-subjects factor. The results revealed a significant difference between the pretest and posttest, $F(1, 18) = 37.26$, $p < .001$, and a significant interaction, $F(1, 18) = 13.62$, $p < .002$. Differences among the means were examined with Tukey post hoc tests. The results of those analyses showed that the improvement between the pretest and posttest scores was significant only for those students who worked with the simulated arms ($p < .001$). Further, there was no significant difference between the two groups on the pretest, but posttest scores for the simulated arm group were statistically higher than those for the CathSim™ group ($p < .01$).

Performance metrics were collected for all of the students in the CathSim™ group during each training session. The means for both sessions are shown in Table 2. The results

indicate that students were able to successfully complete the procedures successfully about 94% to 95% of time across sessions. The mean tourniquet time decreased from session 1 to session 2, $t(8) = 6.59, p < .001$. The overall instances of hematoma also decreased from the first to the second training session. A chi-square test on the frequencies indicated that this drop was statistically significant, $\chi^2(1) = 4.314, p < .05$. Last, the mean pain factor declined slightly from Session 1 to Session 2, but the difference failed to reach significance, $t(8) = 2.18, p = .061$.

Common Steps

During the study, it became clear that there were several important differences between the two training methods. Although both systems enable training on phlebotomy, neither device permits trainees to practice *all* of the steps required to successfully perform the procedure. Further, each device differs in the number of steps it addresses and the method for performing those steps with either device may differ qualitatively from performing the same steps with a genuine patient.

A post hoc analysis was performed to address these differences. The 24 steps for performing phlebotomy assessed in this study are listed in Table 3. Of those 24 steps, all can be practiced with the simulated arm but only 14 can be practiced with the CathSim™ system. For example, both devices allow the user to practice inserting a needle (Step 15); however, only the simulated arm allows the user to practice placing gauze over the needle prior to needle withdrawal (Step 20). Further, those steps marked with an asterisk cannot be performed as they would with a genuine patient. For example, neither system allows the student to actually palpate an insertion site (Step 6). Similarly, although the CathSim™ system requires the user to apply a tourniquet, the step is accomplished using the computer's mouse. The user cannot practice tying the tourniquet around a patient's arm.

Given these differences, it is possible that the students who trained with the simulated arms achieved higher scores because they had the opportunity to practice more steps from the full

procedure. Thus, a second analysis was performed on only those 14 steps common to both training methods. The results of that analysis are shown on the right side of Table 1. As can be seen in the table, the pattern of results is similar to that of the full procedure. Again, there was a significant difference between the pretest and posttest, $F(1, 18) = 28.19$, $p < .001$, and a significant interaction, $F(1, 18) = 6.1$, $p < .025$. Subsequent Tukey tests indicated a significant improvement for only those students who worked with the simulated arms ($p < .001$). Again, there was no significant difference between the two groups on the pretest, but scores for the simulated limb group were statistically higher than for the CathSim™ group on the posttest ($p < .01$).

Specific differences between the two systems on the 14 common steps are shown in Table 4. Mean scores achieved on each step and the proportion of students who correctly performed and received full credit for each step are shown in the table. Although these data were not analyzed statistically due to the potential for inflated alpha levels given the large number of comparisons, some trends are apparent. Performance was comparable on the two systems for 9 of the 14 steps. On the remaining 5 steps (marked with an asterisk in the table), however, there was an advantage for the simulated arm. For these steps, the overall mean scores were greater and a higher proportion of students were able to achieve the maximum score having trained with the simulated arm.

Discussion

The primary goal of the present study was to compare the CathSim™ VR system and simulated limbs for training phlebotomy. As noted earlier, there are no universal standards for training phlebotomists and performance can vary widely from clinic to clinic and person to person. These systems offer the opportunity for students to acquire the skills needed to perform phlebotomy by practicing on devices instead of patients. Although the results showed that all participants benefited to some degree from their training, improvement from the pretest to

posttest was statistically significant only for those students who practiced with the simulated arms. In this regard, these data are consistent with those of other studies showing either no advantage for a VR system over simulated limbs (Chang et al., 2002) or a disadvantage of VR compared to simulated limbs for training IV cannulation (Engum et al., 2003; Scerbo et al., 2004). The results from this study and the others suggest that differences between the cannulation and phlebotomy procedures are not significant enough to produce an advantage for training with the CathSim™ system over the more traditional method using simulated arms.

A Detailed Comparison of the CathSim™ and Simulated Limb Training Devices

Becoming adept at phlebotomy requires trainees to practice all steps of the procedure using equipment and activities that are as similar to the actual task as possible. Doing so ensures that the learning experience is relevant and complete. It also allows for positive transfer from the training situation to the applied task. Classical approaches to training transfer emphasize similarity between the basic elements of the training task and the target task (identical elements theory) or between the principles conveyed by the training and the target tasks (transfer-through-principle theory; Holding, 1965).

Central to the success of any training method is the realism, or fidelity, of the training task. Hays and Singer (1989) explain that fidelity can be expressed both physically and functionally. These two types of fidelity may be mapped directly onto the conditions of "stimulus" and "response" similarity discussed by Holding (1965). Training systems that have high physical fidelity include many of the same sensory stimuli that are part of the actual equipment. Those that have high functional fidelity require many of the same responses or actions required by the actual equipment. Ideally, an optimal training system would have both high physical and functional fidelity.

Specifying physical and functional fidelity differences between the 14 common steps trained by the CathSim™ system and the simulated limb may help clarify why the participants who trained with the simulated limb performed better on the posttest.

Regarding physical fidelity, the two devices vary most on the following steps: tourniquet application (the simulated limb uses an actual tourniquet - CathSim™ displays an icon); preparing the site with alcohol or betadine (simulated limb training uses an actual alcohol pad for better monitoring of application technique - CathSim™ displays an icon); stabilizing the vein (the simulated limb allows participants to experience feeling a vein with their fingers - CathSim™ has a flat rubber pad); and orienting the needle properly (the simulated limb uses a genuine needle so that one can monitor the orientation of the beveled end - CathSim™ uses a facsimile of a needle without beveled end).

With respect to functional fidelity, there are also differences between the two devices. This is of particular relevance, as some researchers have argued that functional fidelity may be more important than physical fidelity for ensuring acquisition of skills (Gopher, Weil, Bareket, & Caspi, 1988). Because an actual tourniquet was used with the simulated limb the trainees were able to practice tying the tourniquet. On the CathSim™ system, the tourniquet is placed by clicking the tourniquet icon. This functional difference may have contributed to the lower posttest scores of CathSim™ students on this step. Although both devices allow participants to cleanse the site with alcohol, only the simulated limb permits the opportunity to confirm that the site is dry before proceeding. This difference may have led to lower posttest scores for the CathSim™ students on this step as well. In addition, the simulated limb provides some opportunity to practice stabilizing the vein (i.e., students can position their fingers on the vein, but the vein is largely fixed and does not move much). On the CathSim™ system, however, this step must be performed on the simulated skin pad of the AcuTouch™ device which is flat and does not have any protruding areas to represent veins. This is another functional difference that may have led to lower posttest scores for the CathSim™ students on this step. Finally, as noted earlier ensuring

that the needle enters the site bevel-side-up requires more proactive effort for participants using the simulated limb. This step cannot be performed properly with the CathSim™ system and may also have contributed to the lowers posttest scores for these students.

In addition to these differences, there are also some sensory deficiencies inherent in CathSim™ system design. Visually, it is not possible to monitor all aspects of the phlebotomy procedure because of the system's limited field of vision. Auditorially, CathSim™ provides utterances to signify the experience of pain by the patient; however, there are no other auditory signals provided such as heart rate, respiration, or patient commentary. There are two other design-oriented concerns regarding the CathSim™ system. First, the CathSim™ trainer is most easily used when seated, whereas phlebotomy is typically performed while standing. This difference influences other task performance steps, including the angle of needle insertion. Also, because CathSim™ users must interact simultaneously with the computer mouse and the needle/hub assembly, there is potential for motor interference.

The physical and functional differences between the two systems may help to explain why the two groups attained different scores on the posttest and also provide guidance for designers of virtual phlebotomy systems and other medical training devices. The performance differences observed in the present study suggest that limitations in virtual simulators that compromise physical and functional fidelity can undermine the very training benefits espoused by proponents of this technology.

Despite these design limitations, the CathSim™ system does provide other learning experiences. In particular, the system records a variety of performance metrics that are not available or easily obtainable with simulated limbs. For example, students are typically instructed not to keep a tourniquet fastened around a patient's arm for longer than two minutes because it can affect the quality of the blood sample. As noted in Table 2, the mean tourniquet time decreased from about 99 to 73 seconds across training sessions. However, there were 18 instances in the first training session where the tourniquet time exceeded the 2-minute limit and

the mean of those 18 instances was 180 seconds. By the second session, there were only two instances that exceeded the limit and the mean was 137 seconds. If these attempts had been conducted with genuine patients, the procedure would have needed to be terminated and started on another arm. The CathSim™ system also records an index of hematoma that is not observable with the simulated limbs. As noted above, there was a significant decrease in occurrences of hematoma from the first to the second training session. Thus, by practicing on the simulator one could argue that 23 instances of potential hematoma were prevented from occurring in actual patients. This is an important point because performance in the present study was assessed on the simulated limbs and not genuine patients. Thus, the CathSim™ system may provide a critical measure of performance that would be observable in actual patients, but not assessed with the simulated limbs used for the pretest and posttest in the present study.

In addition, the CathSim™ system also provides exposure to a limited range of patient characteristics. Further, the system includes two auxiliary visual aids (a cut-away side view that shows the needle penetrating the arm and a transparency view of the underlying vascular system) that were not examined in the present study.

Although the functional and physical differences between the two training systems may explain the lower posttest scores for the CathSim™ students, it is important to remember that neither system completely or faithfully represents the phlebotomy task as it is performed in practice. As noted in Table 3, neither device allows trainees to practice palpating the insertion site nor re-inspecting the site for bleeding as one would do with a genuine patient. Also, site selection posed no real challenge. The veins in the simulated limb protrude from the arm making site selection obvious. Further, the skin covering the simulated limb retains puncture marks that subsequent trainees can see and use for guidance. On CathSim™, the computer's mouse is used to select a site and will only allow palpation in specific areas of the arm. Further, neither system permits the user to "damage" nearby structures and therefore students do not have an opportunity to learn about the risks associated with poorly executed procedures. Finally, neither the simulated

limb nor the CathSim™ system incorporates the stress, workload, or level of feedback present in an actual clinical setting.

In sum, there may be advantages and disadvantages to both the CathSim™ system and simulated limbs for training phlebotomy. In fact, the optimal training program may be one that incorporates both systems and capitalizes on the unique advantages afforded by each. Moreover, the opportunity to practice on either system before attempting the procedure on genuine patients could potentially decreases risks to patients as well as the incidence of needlestick injuries across all healthcare facilities.

Future Research Issues

The results from the present study suggest that simulated limbs may be more effective than a VR system for training phlebotomy. However, it is important to understand that *both* groups were assessed on the simulated limbs. Thus, one could argue that the students who trained with the CathSim™ system were at a disadvantage because they were assessed on a system that differed from the one on which they practiced. A decision was made to use simulated limbs for the pretest and posttest in the present study for several reasons. First, we wanted to obtain a baseline measure of performance on as many steps of the procedure as possible and for obvious safety reasons, we could not permit inexperienced students to draw blood from genuine patients without supervision or the possibility of having a supervisor intervene. Second, we wanted the baseline measure to address performance on the medical procedure and therefore be free from activities needed to familiarize oneself with the computer interface for the VR system. Last, the simulated limbs offered a standardized platform for assessment. Although this approach enables comparisons of performance before and after training, a more genuine and clearer picture of the efficacy of the CathSim™ and simulated limb methods necessitates that performance be

evaluated on a "neutral" target task. Ultimately, these training methods must be evaluated where it counts most -- with genuine patients.

To date, most of the data supporting the benefits of VR medical simulators have not been gathered from actual patients. Instead, they have been generated from the systems themselves. There are numerous technical and practical issues that impact efforts to evaluate medical simulators in clinical settings (see Satava & Jones, 2002; Scerbo, in press); however, progress is being made. Recently, Seymour and his colleagues (2002) reported some training benefits with a VR system for fundamental laparoscopic skills gathered from real patients.

Although evaluations based on system-generated metrics of the sort reported in this study are valuable, they are limited to demonstrating improvement over time and distinguishing between expert and novice performance. However, true measures of training transfer can be derived only from data collected in the target task environment (i.e., with real patients). Moreover, the ability to assess the predictive validity of system-generated metrics and to investigate the concurrent validity of simulator-based training with other procedures all hinge upon gathering data with genuine patients. Accordingly, to address this limitation of the present study a second experiment is currently under way to examine the degree of training transfer afforded by the CathSim™ and simulated limb methods for phlebotomy with genuine patients.

The results from studies such as ours are a necessary first step to help establish the effectiveness of VR medical simulators if the science and practice of medicine is to benefit from this technology and evolve. Ultimately, however, we need to understand the degree to which time spent on simulators predicts performance with genuine patients so that we can study the science and practice of medicine in a safe, controlled environment. Just as flight simulators allow researchers to examine pilot performance under a wide variety of conditions, medical simulators promise to allow the study of physicians, surgeons, and other medical personnel under an unlimited array of contexts. For example, simulators may be used to study the effects of fatigue, stress, workload, and work schedules on surgical skills, decision-making, and problem solving.

More important, research on potential moderating factors and countermeasures can also be studied. Medical VR simulators can help establish more efficient methods for training by capitalizing on principles of knowledge and skill acquisition that address task sequencing, part and whole task training, and integrative approaches (Proctor & Dutta, 1995); the development of automatic skills (Schneider, 1985); and adaptive training paradigms (Krahl & Scerbo, 1997). Medical VR simulators can also be targeted to individuals or teams of surgeons, residents, anesthesiologists, and nurses (Helmreich & Schaefer, 1994).

Conclusion

Recently, the Institute of Medicine issued a report, *To Err is Human* (Kohn, Corrigan, & Donaldson, 1999) that suggested medical errors contributed to as many as 98,000 deaths annually in U.S. hospitals. A more recent report issued by Health Grades, Inc. (2004) suggests that the number of fatalities may be double the original estimate. In response to concerns about resident fatigue and patient safety, the American Medical Association's Accreditation Council for Graduate Medical Education, the organization that accredits teaching hospitals in the U.S., recently set limits on the working hours for medical residents (2003). Although residents may welcome the reduction in working hours, the restrictions have created new pressures to train physicians and surgeons to higher levels of competency in shorter periods of time.

Medical VR simulators may provide a partial solution to these problems. They offer the opportunity for students, residents, and even practicing physicians to have access to alternative forms of training at any time of the day or night. Despite this obvious advantage, considerable work still needs to be done to validate the systems and establish rates of training transfer. The results from the present study indicate that a medical VR system for phlebotomy may not be as effective as a more traditional low-tech method due to functional and physical differences between the systems. Clearly, there is a need for those in the human factors community to

participate in the design and evaluation of medical VR simulators so that this technology can evolve with the greatest benefits for the end user.

Acknowledgements

This study was a collaborative project between the Virginia Modeling, Analysis and Simulation Center (VMASC) at Old Dominion University and the Eastern Virginia Medical School. Partial funding for this study was provided by the Telemedicine and Advanced Technologies Research Center, U.S. Army Medical Research and Material Command, contract DAMD17-03-2-0004 and the Naval Health Research Center through NAVAIR Orlando TSD under contract N61339-03-C-0157, entitled "The National Center for Collaboration in Medical Modeling and Simulation". The ideas and opinions presented in this paper represent the views of the authors and do not necessarily represent the views of the Department of Defense.

References

American Medical Association (2003). AMA delegates approve limits on resident working hours. Retrieved August 12, 2003, from http://www.ama-assn.org/ama/pub/article/1616-6390.html.

Basdogan, C., & Srinivasan, M. A. (2002). Haptic rendering in virtual environments. In K. M. Stanney (Ed.), *Handbook of virtual environments: design, implementation and applications* (pp. 117-134). Mahwah, NJ: Erlbaum.

Chang, K. K., Chung, J. W., & Wong, T. K. (2002). Learning intravenous cannulation: A comparison of the conventional method and the CathSim™ intravenous training system. *Journal of Clinical Nursing, 11*, 73-78.

Clover, S. (2002). Phlebotomy safety: Sticking to good practices. Advance Newsmagazines for Medical Laboratory Professionals, July. ONLINE: http://www.advanceformlp.com/common/editorial/PrintFriendly.aspx?CC=6475.

Dawson, S. L. (2002). A critical approach to medical simulation. *Bulletin of the American College of Surgeons, 87*(11), 12-18.

Engum, S.A., Jeffries, P., & Fisher, L. (2003). Intravenous catheter training system: computer-based education versus traditional learning methods. *American Journal of Surgery, 186,* 67-74.

Gopher, D., Weil, M., Bareket, T., & Caspi, S. (1988). Fidelity of task structure as a guiding principle in the development of skill trainers based upon complex computer games. Proceedings of the 32nd Annual Human Factors Society Annual Meeting. Santa Monica, CA: Human Factors Society.

Hays, R.T., & Singer, M.J. (1989). *Simulator fidelity in training system design.* New York: Springer-Verlag.

Health Grades, Inc. (2004). *Patient safety in American hospitals.* Lakewood, CO: Author.

Helmreich, R. L., & Schaefer, H. (1994). Team performance in the operating room. In M. S. Bogner (Ed.), *Human error in medicine* (pp. 225-253). Hillsdale, NJ: Lawrence Erlbaum.

Holding, D.H. (1965). *Principles of training.* London: Pergamon Press.

Howanitz, P. J., Cembrowski, G. S., & Bachner, P. (1991). Laboratory phlebotomy: College of American Pathologists Q-Probe study of patient satisfaction and complications in 23 783 patients. *Archives of Pathology and Laboratory Medicine, 115*, 867-872.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S., Eds. (1999). *To err is human: Building a safer health system.* Washington, D.C.: National Academy Press.

Krahl, K., & Scerbo, M.W. (1997). Performance on an adaptive automated team tracking task with human and computer teammates. *Proceedings of the Human Factors & Ergonomics Society 41st Annual Meeting* (pp. 551-555). Santa Monica, CA: Human Factors & Ergonomics Society.

Mishori, R. (2004, June 1). Drawing without a license: Phlebotomists get little training, regulation. *The Washington Post,* p. HE1.

NIOSH. (1999). How to Protect Yourself From Needlestick Injuries. Retrieved June 9, 2004, from http://www.cdc.gov/niosh/2000-135.html.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance.* Thousand Oaks, CA: Sage Publications.

Prystowsky, J.B., Regehr, G., Rogers, D.A., Loan, J.P., Hiemenz, L.L., & Smith, K.M. (1999). A virtual reality module for intravenous catheter placement. *The American Journal of Surgery, 177,* 171-175.

Reznek, M. A., Rawn, C. L., Krummel, T. M. (2002). Evaluation of the educational effectiveness of a virtual reality intravenous insertion simulator. *Academic Emergency Medicine, 9,* 1319-1325.

Satava, R. M. (1993). Virtual reality surgical simulator: The first steps. *Surgical Endoscopy, 7,* 203-205

Satava, R. M. (2001). Accomplishments and challenges of surgical simulation: Dawning of the next-generation surgical education. *Surgical Endoscopy, 15,* 232-241.

Satava, R. M., & Jones, S. B. (2002). Medical applications of virtual environments. In K. M. Stanney (Ed.), *Handbook of virtual environments: design, implementation and applications* (pp. 937-957). Mahwah, NJ: Erlbaum.

Scerbo, M. W. (in press). Medical virtual reality simulators. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors, 2nd Ed.* (pp.). CRC Press.

Scerbo, M. W., Bliss, J. P., Schmidt, E. A., Thompson, S. N., Cox, T. D., & Poland, H. J. (2004). A Comparison of the CathSim™ system and simulated limbs for teaching intravenous cannulation. In J. D. Westwood et al. (Eds.), *Medicine meets virtual reality, 12,* (196-198). Amsterdam: IOS Press.

Schneider, W. (1985). Training high-performance skills: fallacies and guidelines. *Human Factors, 27,* 285-300.

Seymour, N. E., Gallagher, A. G., Roman, A. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. *Annals of Surgery, 236,* 458-464.

Ursino, M., Tasto, J. L., Nguyen, B.H., Cunningham, R., & Merril, G. L. (1999). CathSim™: An intravenous catheterization simulator on a PC. In J. D. Westwood et al. (Eds.), *Medicine meets virtual reality,* (360-366). Amsterdam: IOS Press.

Author Note

Address correspondence to Mark W. Scerbo, Department of Psychology, Old Dominion University, Norfolk, VA 23529-0267; mscerbo@odu.edu.

Table 1

Mean Assessment Scores on the Pretest and Posttest for Both Groups on the Full Procedure and on Common Steps

	Full Procedure		Common Steps	
	Pretest	Posttest	Pretest	Posttest
Simulated Arms	60.82 (8.6)	82.91 (3.5)	36.91 (6.5)	49.1 (2.3)
VR System	63.22 (13.8)	68.67 (7.2)	37.78 (8.6)	42.22 (5.3)

(Standard deviations in parentheses)

Table 2

Performance Metrics for the VR System Participants on Each Training Session Collapsed Across Trials and Cases

Metric	Session 1	Session 2
Percent of Successful Procedures	94.1	95.4
Mean Tourniquet Time	98.65 (12.33)	72.9 (14)
Frequency of Tourniquet Times Exceeding Two Minutes	18	2
Frequency of Hematoma		
Moderate	10	2
Serious	8	3
Mean Pain Score	3.05 (1)	2.18 (.92)

(Standard deviations in parentheses)

Table 3

Steps of the Phlebotomy Procedure that can be Practiced Using the Simulated Arm (S) and CathSim™ System (C)

S		1. Gather and arrange all materials (tourniquet, Vacutainers, gauze pads, etc.)
S		2. Wash hands
S		3. Position tourniquet
S	C*	4. Apply tourniquet
S*	C	5. Select phlebotomy site
S*	C*	6. Palpate venipunture site
S		7. Put on gloves
S	C*	8. Prepare site with alcohol or beta iodine stick
S		9. Sterilize collection tube tops with alcohol pad
S	C	10. Grasp needle holder behind the needle hub
S		11. Remove the needle's protective sleeve
S	C*	12. Stabilize vein
S	C	13. Hold needle at 15-30 degree angle
S	C*	14. Enter site bevel side up
S	C	15. Insert needle
S	C	16. Advance needle into vein
S	C	17. Depress tubes into Vacutainer until filling stops
S	C	18. Disengage collection tube
S	C*	19. Release tourniquet
S		20. Place gauze pad over needle
S	C	21. Withdraw needle
S		22. Press gauze pad onto puncture site and apply pressure
S		23. Properly label specimens
S*		24. Recheck puncture site for bleeding

* indicates limited or partial ability to practice the step

Table 4

Mean Scores and Proportions of Students who Correctly Performed the Fourteen Common Steps of the Phlebotomy Procedure Using the Simulated Arm (S) and CathSim™ System (C)

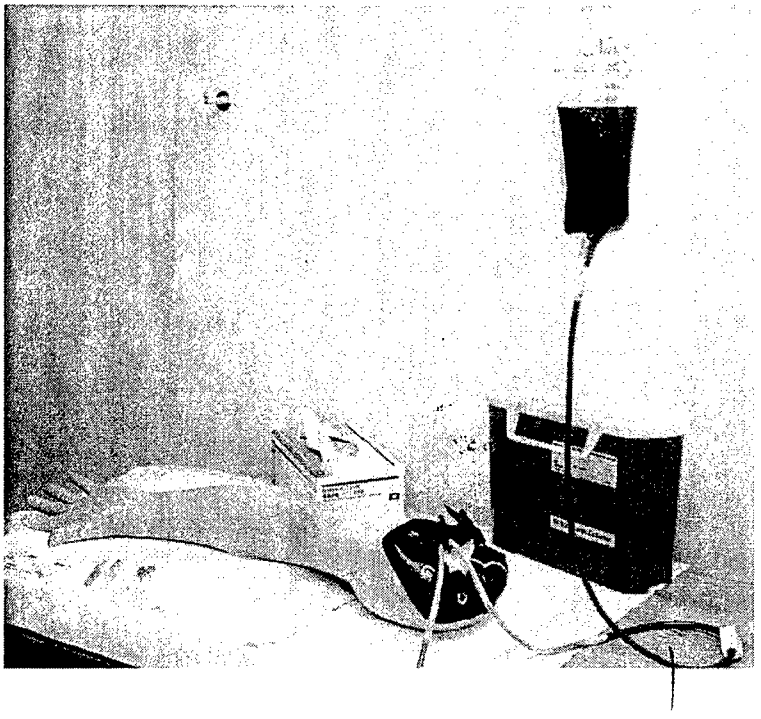
Common Steps	Limited Fidelity	Sim. Arm Mean Score	CathSim™ Mean Score	Sim. Arm Proportion Correct	CathSim™ Proportion Correct
Apply tourniquet *	C	4.44	2.67	.82	.33
Select site	S	4.00	3.11	1.00	.78
Palpate site	S, C	2.00	1.77	1.00	.89
Prepare site *	C	4.00	2.67	1.00	.44
Grasp needle holder *		2.00	1.11	1.00	.56
Stabilize vein *	C	3.64	0.67	.82	.11
Angle needle properly		4.00	3.77	1.00	.89
Enter site, bevel up *	C	4.00	3.33	1.00	.67
Insert Needle		3.82	3.77	.91	.89
Advance needle		6.00	6.00	1.00	1.00
Depress tubes		3.27	3.55	.73	.78
Disengage tube		4.00	4.00	1.00	1.00
Release tourniquet	C	1.82	2.00	.91	1.00
Withdraw needle		3.82	3.33	.91	.78

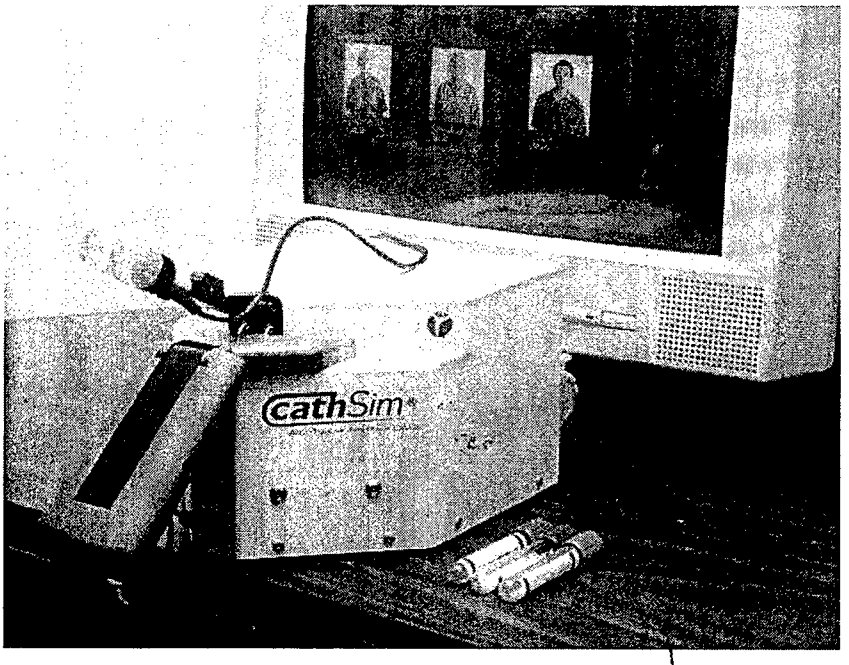
* indicates an advantage for the simulated arm over the CathSim™ system

Figure Captions

Figure 1. The Life/Form® simulated arm with standard IV bag brings artificial blood into the tubing of the arm.

Figure 2. The CathSim™ VR multimedia system with AccuTouch™ six-degree-of-freedom haptic feedback device that simulates the needle and hub assembly and includes a vinyl strip to simulates the skin.




CathSim

Author Biographies

MARK W. SCERBO received his Ph.D. from the University of Cincinnati in 1987. He is a professor at Old Dominion University. His research interests include modeling and simulation in medicine and other domains, adaptive automation, vigilance and attention, and visual perception.

JAMES P. BLISS received his Ph.D. from the University of Central Florida in 1993 and is an associate professor at Old Dominion University. He currently conducts research in the areas of visual and auditory alarm mistrust, modeling and simulation, and training.

ELIZABETH A. SCHMIDT is a doctoral student at Old Dominion University. Her current research interests are in the area of medical modeling, simulation, and training.

SOMMER N. THOMPSON received her B. S. with honors in psychology from Old Dominion University in 2004. She is interested in pursuing her doctoral degree in human factors psychology.

A.3 Elizabeth A. Schmidt, Mark W. Scerbo, James P. Bliss, and Sommer N. Thompson, "Skill Acquisition with a VR Simulator for Phlebotomy," *Second Conference on Human Performance, Situation Awareness, and Automation*, Daytona Beach, FL., Mar., 2004.

Presented at the 2nd Conference on Human Performance, Situation Awareness, and Automation, Daytona Beach, FL., Mar., 2004.

Skill Acquisition with a VR Simulator for Phlebotomy

Elizabeth A. Schmidt, Mark W. Scerbo, James P. Bliss, and Sommer N. Thompson
Old Dominion University

ABSTRACT

The present study was designed to examine skill acquisition using a VR simulator for phlebotomy. Participants were given five hours of individualized training on the VR system. Half of the participants were allowed use two visual aids and the remaining participants practiced without the aids. The results showed no benefits for the visual aids and performance for all participants became fairly stable after approximately 45 minutes of practice. Data from pretest and posttest scores also showed improvement, but the gains were not dramatic. The results from this study and others suggest that characteristics of the VR system may limit the degree to which skills acquired on the system transfer to other contexts.

INTRODUCTION

The use of high fidelity virtual reality (VR) simulators has had a profound impact on operator performance in aerospace and the military for decades. In medicine, the first virtual reality simulators began to appear in the early 1990s (Satava, 1993). Since then, there has been a proliferation of systems to address a variety of procedures. The traditional approach for learning medical procedures has been to follow an apprenticeship model and "see one, do one, teach one." However, with the advent of VR technology, medical training can be performed in conjunction with immersive simulators. The use of immersive simulators as an alternative to the apprenticeship model allows trainees to practice in a safe environment without the need for patients or animal labs. However, use of the VR technology to learn medical skills has not been widely tested and validated (Magee, 2003; Satava, 2001).

The Present Study

The present study was designed to examine skill acquisition using an immersive simulator for phlebotomy. The objective was to determine if the skills required for performing phlebotomy could be acquired by training on the CathSim™ simulator. In addition, two visual performance aids were studied to determine their effect on skill acquisition. It was expected that performance would improve as a result of continued practice and that the visual aids would have a positive effect on the rate of learning.

METHOD

Participants

Participants were 10 graduate students in a psychology program at Old Dominion University. They ranged in age from 22 to 32 ($M = 25.8$, $SD = 3.4$). Only one participant had any experience with phlebotomy, reporting minimal experience drawing blood from animals. All participants were volunteers who received monetary compensation for their participation.

Materials

Simulated arm. A Life/Form® simulated arm was used for the pretest and posttest. The simulated arm has a layer of vinyl skin covering a network of latex veins (see Figure 1). Students' phlebotomy skills using the simulated arm were assessed before and after VR training using a standard assessment form, allowing performance assessment on a scale from 0-88.

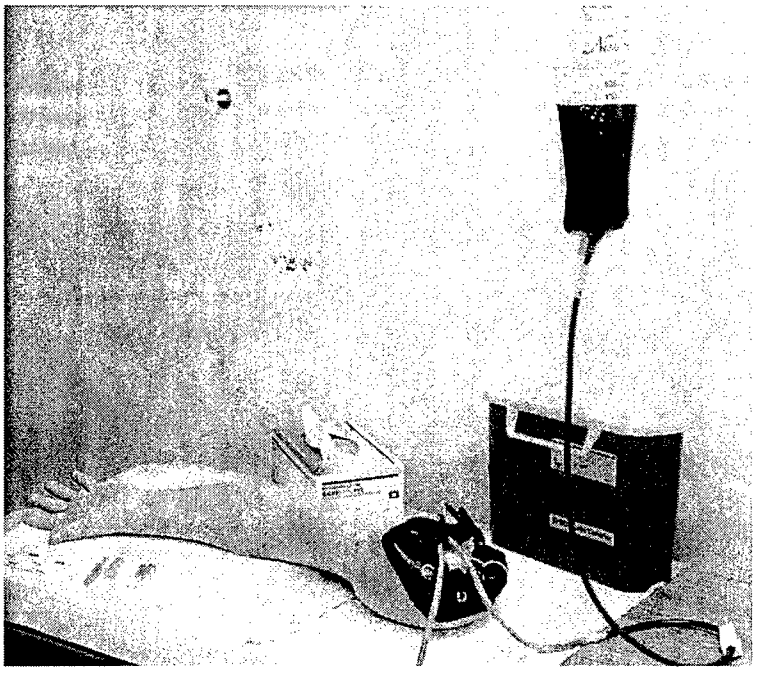


Figure 1. The Life/Form® simulated arm.

CathSim™ system. The VR simulator used in the present study was the CathSim™ system from Immersion Medical, Inc. (see Figure 2). CathSim™ simulates phlebotomy procedures using visual, auditory, and haptic displays. The physical system consists of an IBM-compatible computer accompanied by an AccuTouch™ six-degree-of-freedom haptic feedback device that simulates the needle and a vinyl strip that simulates the skin.

Procedure

All participants watched an instructional video about phlebotomy and then performed a pretest phlebotomy on the simulated arm. They were each given five hours of individualized training with all six case patients provided with the CathSim™ system. All participants were required to meet several criteria (i.e., perform a successful procedure with no hematoma, a tourniquet time of less than 90 seconds, and a low pain score) on each case during their training sessions. Once training was complete, all participants performed a posttest phlebotomy on the simulated limb. Half of the participants were allowed use two visual aids: a cut-away side view that showed the needle penetrating the arm and a transparency view of the underlying vascular system. The remaining participants practiced without the visual aids.

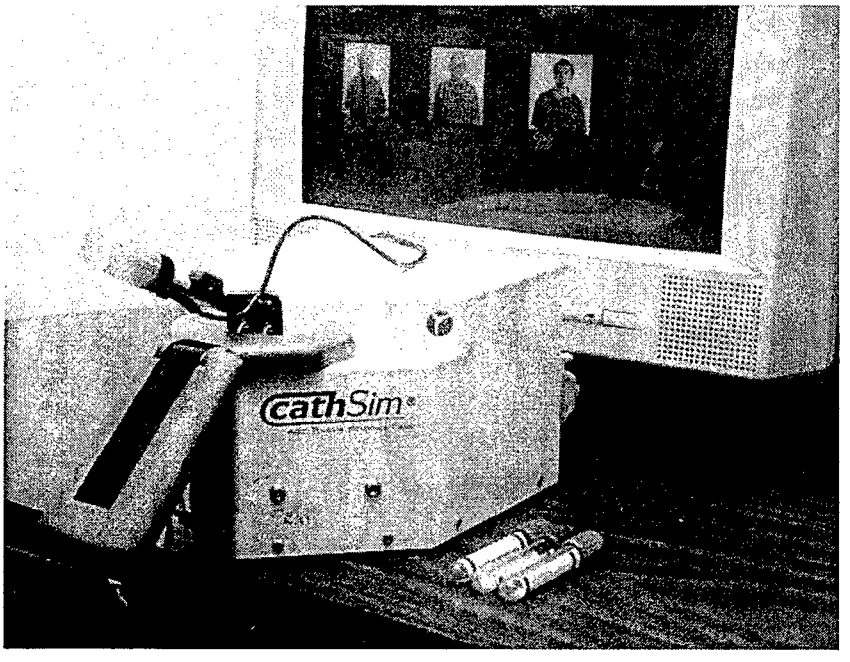


Figure 2. The CathSim™ VR system.

RESULTS

Pretest-Posttest

One individual did not complete the training, so the results are based upon nine participants. Further, initial analyses revealed no effects of the visual aids, so those data were collapsed across groups. Pretest and posttest scores were compared using a dependent t-test. Performance improved from pretest (M=48.56, SD=10.24) to posttest (M=63.89, SD=10.13), $t(8)$ <.01. Although the tourniquet times (an index of performance efficiency) decreased from pretest to posttest, the decline was not significant.

Learning Curves

Skill acquisition over trials (case patients) is shown in Figure 3. The data show the mean tourniquet times obtained from each participant for each case patient. Each participant required a different number of attempts to reach criterion on each case patient. The total times needed to reach criterion for all cases were collapsed across participants and are shown in Figure 4. As can be seen in the figure, most of the performance gain was achieved across the first 3 trials. Beyond that, continued practice resulted in small performance benefits.

DISCUSSION

The present study was designed to examine how visual aids facilitate skill acquisition with a VR simulator for phlebotomy. The results showed no appreciable performance benefit for providing users with a cut-away side view or transparent view of the underlying vascular system. Participants in both groups learned at the same rate and scored similarly on the posttest.

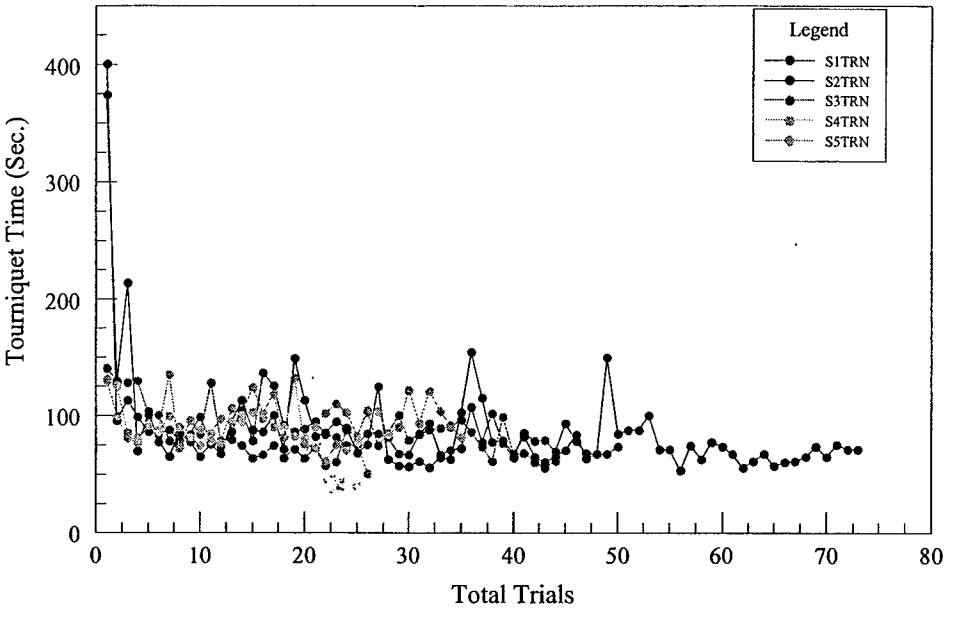


Figure 3. Mean tourniquet times as a function of trials (case patients). Data from each participant are represented by different colors.

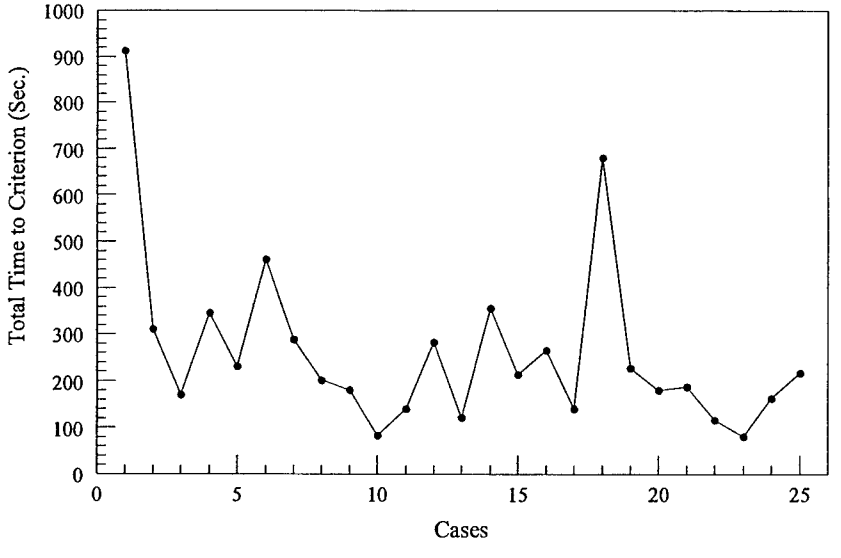


Figure 4. Total time needed to reach criterion collapsed across participants.

An examination of the learning curves showed that performance became fairly stable after only a few cases (approximately 45 minutes). However, performance remained variable across the entire training session. It is possible that much of this variability could be attributed to the different levels of difficulty associated with each case patient.

Overall, participants showed improvement between the pretest and posttest, but the gains were not dramatic. Posttest scores were 30% higher than pretest scores after five hours of practice. In another study of the CathSim™ system addressing intravenous catheterization, Scerbo and his colleagues (2004) found that the VR system was a less effective training method than simulated limbs. They argued that the CathSim™ system as compared to simulated limbs provides the opportunity to practice fewer steps from the full procedure. Further, the activities required to perform several steps on the CathSim™ system did not faithfully reproduce the activities necessary for implementing those steps on a simulated arm or on a genuine patient. Thus, it is possible that the meager gains in performance observed on the posttest in the present study suggest that skills acquired on the CathSim™ system do readily transfer to other contexts, either simulated limbs or even genuine patients.

REFERENCES

Magee, J. H. (2003). Validation of medical modeling & simulation training devices and systems. In J. D. Westwood et al. (Eds.), *Medicine meets virtual reality, 11*, (196-198). Amsterdam: IOS Press.

Satava, R. M. (1993). Virtual reality surgical simulator: The first steps. *Surgical Endoscopy, 7*, 203-205.

Satava, R. M. (2001). Accomplishments and challenges of surgical simulation: Dawning of the next-generation surgical education. *Surgical Endoscopy, 15*, 232-241.

Scerbo, M. W., Bliss, J. P., Schmidt, E. A., Thompson, S. N., Cox, T. D., & Poland, H. J. (2004). A Comparison of the CathSim™ system and simulated limbs for teaching intravenous cannulation. In J. D. Westwood et al. (Eds.), *Medicine meets virtual reality, 12*, (196-198). Amsterdam: IOS Press.

A.4 James P. Bliss, Hope S. Hanner-Bailey, & Mark W. Scerbo, "Determining the Efficacy of an Immersive Trainer for Arthroscopy Skills," presented at *Medicine Meets Virtual Reality XIII*, Long Beach, CA, Jan., 2005.

To be presented at Medicine Meets Virtual Reality XIII, Long Beach, CA, Jan., 2005.

Determining the Efficacy of an Immersive Trainer for Arthroscopy Skills

James P. BLISS, Hope S. HANNER-BAILEY, & Mark W. SCERBO
Old Dominion University

Abstract. The present study examined the effectiveness of an immersive arthroscopic simulator for training naïve participants to identify major anatomical structures and manipulate the arthroscope and probe. Ten psychology graduate students engaged in five consecutive days of practice sessions with the arthroscopic trainer. Following each session, participants were tested to see how quickly and accurately they could identify 10 anatomical landmarks and manipulate the arthroscope and probe. The results demonstrated steady learning on both tasks. For the anatomy task, participants correctly identified an average of 7.7 out of 10 structures correctly in the first session and 9.5 in the last. During the manipulation task, participants collided 53.5 times with simulated tissues in the first session and 13.2 times during the final session. Participants (n=9) also demonstrated minimal performance degradation when tested 4 weeks later. These data suggest that the immersive arthroscopic trainer might be useful as an initial screening or training tool for beginning medical students.

1. Background

The knee has been characterized as the most complex and frequently injured joint in the human body [1]. Knee arthroscopy has been performed for many years, since Kenji Takagi of Tokyo University performed the procedure first in 1918. The Virtual Reality in Medicine and Biology Group at the University of Sheffield has developed a VE-based simulator to represent arthroscopy [2], as has the Fraunhofer Institute for Computer Graphics [3]. Some simulators have been designed with force feedback to allow trainees to interact more realistically with anatomical structures of the knee. The increasing popularity and sophistication of arthroscopic simulation has prompted some medical personnel to call for the routine use of VE-based simulation in arthroscopic training [4]. Such training, it is claimed, allows trainees to practice without relying on animal models, cadavers, or actual patients. The current experiment is part of a larger program to investigate the effectiveness of an immersive arthroscopic trainer. Our plan was to determine whether naïve participants could use the simulator to learn fundamental anatomy and probe manipulation and navigation. Based on prior data from other medical simulators, we expected that participants would quickly and successfully demonstrate learning and would retain the acquired knowledge across a four-week period.

2. Methodology

2.1 Participants

Ten psychology graduate students from Old Dominion University participated. It was impractical to use medical students or residents because the present study required

participants to practice with the simulator for five consecutive days and to be tested four weeks later. Participants included four males and six females, and participants did not have prior medical experience. They ranged in age from 21 to 55 years (M=27.5, SD=10.0). The participants were paid $150 for their participation.

2.2 Apparatus and Procedure

The simulator used for this experiment was the Procedicus^{TM} Virtual Arthroscopy (VA) trainer, manufactured by Mentice Corporation. Though the simulator includes modules for training complex skills such as loose body removal and subacromial decompression, we concentrated on only two basic skills: identifying anatomical landmarks and manipulating the arthroscope and probe. Participants were personally contacted to determine their interest in the project. Twenty-four hours before the experiment, participants studied a short information sheet that described ten anatomical landmarks within the knee and several anatomical terms. After arriving at the hospital laboratory, participants completed an informed consent form and were given ten minutes to review the background study material. They then spent 15 minutes manipulating the arthroscope to review knee landmarks. They were then tested to determine their ability to recognize the landmarks. After a brief break, participants spent 15 minutes practicing arthroscope and probe manipulation by locating and intersecting blue balls randomly placed within the simulated knee. After the practice session, participants were tested to see how quickly and accurately they could intersect 11 balls. They then completed an opinion questionnaire and were dismissed until the next day. Participants repeated the experimental procedure for five consecutive days. They were then brought back into the hospital laboratory four weeks later, at which time they repeated the test procedures.

3. Results

Participants demonstrated steady learning on both tasks, correctly identifying an average of 7.7 out of 10.0 anatomy structures in the first session, 8.2 in the second, 8.9 in the third, 9.2 in the fourth, and 9.5 in the last. Manipulation score averages (out of 100) were 2.9, 28.5, 49.1, 54.8, and 56.2 for sessions 1-5, respectively. Mean times to complete manipulation were 495.73, 322.45, 185.33, 167.95, and 161.47 seconds for sessions 1-5 respectively. During the manipulation task, participants collided with simulated tissue 53.5, 25.5, 7.4, 7.7, and 13.2 times, respectively. During the retention session, participants correctly identified an average of 8.4 anatomy structures. Furthermore, the mean manipulation time to complete the manipulation task was 167.31 seconds and the mean number of tissue collisions was 9.0.

In addition to mean differences across time, several significant correlations were observed. Although sex was not found to significantly correlate with any variable, older participants had more video game experience (r = .85, p < .01), and longer retention manipulation times (r = .83, p < .01). Furthermore, participants who reported difficulty with the simulator tended to have lower manipulation scores (r = -.84, p < .01) and more frequent collisions (r = .81, p < .01). Manipulation time was also negatively related to perceived helpfulness of visual aids (r = -.84, p < .01). Questionnaire data indicated that

participants believed that the simulator was an effective tool for learning anatomy and manipulation basics.

4. Discussion

The data reported here illustrate learning rates for naïve participants. Throughout the five days of training, participants performed better and experienced less frustration on the anatomy exercise than they did the manipulation task. The anatomy task required the manipulation of only the scope, while the manipulation exercise required simultaneous use of the scope and probe. Thus, the manipulation task necessitated greater psychomotor ability. Equally important, however, was the level of retention demonstrated. Although there was a slight drop in the number of anatomical structures correctly identified after the four-week period, participants actually made fewer collisions with tissues during retention than they did at Day 5. The time to complete the manipulation task during retention was only marginally longer (5.84 seconds) than it was during the final day of training. Interestingly, there was minimal degradation of performance for both the anatomy and manipulation tasks. The anatomy task may have been slightly more difficult for participants during retention given that it is a more cognitively demanding exercise.

The major drawback of the present study was the small sample size. In the future, researchers should examine the learning curve of the Procedicus VA simulator with a larger sample and across a longer time period. Despite these limitations, it was encouraging that the tasks were not too difficult for naïve trainees to master. This suggests that the arthroscopic trainer might be useful as an initial screening or training tool for beginning medical students.

5. Conclusions

These data show that naïve participants can use the Procedicus VA trainer to learn basic anatomy and tool manipulation skills. Of the two tasks, manipulation seems to be more difficult, and may require extended practice for participants to master it.

References

[1] McCarthy, A. D., & Hollands, R. J. (1997). Human factors related to a virtual reality surgical simulator: The Sheffield Knee Arthroscopic Training System. *Proceedings of the 4th Virtual Reality Special Interest Group Conference.* Leicester, England.

[2] Hollands, R. J., Trowbridge, E. A., Bickerstaff, D., Edwards, J. B., & Mort, N. (1995). The particular problem of arthroscopic surgical simulation: A preliminary report. In R. A. Earnshaw & J. A. Vince (Eds.) Computer *Graphics: Development in Virtual Environments.* London: Academic Press.

[3] Ziegler, R., Brandt, C., Kunstmann, C., Mueller, W., & Werkhaeuser, H. (1997). Haptic displays for the VR arthroscopy simulator. *Proceedings of the SPIE Conference.*

[4] Hardy, P. (2001). Arthroscopic surgery and virtual reality : Development of a virtual arthroscopic trainer. Newsletter of the International Society of Arthroscopy, Knee Surgery, and O Paper presented at the Advanced Technology Applications for Combat Casualty Care Annual Meeting, St. Pete Beach, FL.

A.5 Mark W. Bowyer, Elisabeth A. Pimentel, Jennifer B. Fellows, Ryan L. Scofield, Vincent L. Ackerman, Patrick E. Horne, Alan V. Liu, Gerald R. Schwartz, and Mark W. Scerbo, "Teaching Intravenous Cannulation to Medical Students: Comparative Analysis of Two Simulators and Two Traditional Educational Approaches," In J.D. Westwood et. al. (Eds.), *Medicine Meets Virtual Reality*, 13, (57-63). Amsterdam: IOS Press, 2005.

Medicine Meets Virtual Reality 13
James D. Westwood et al. (Eds.)
IOS Press, 2005

Teaching Intravenous Cannulation to Medical Students: Comparative Analysis of Two Simulators and Two Traditional Educational Approaches

Mark W. BOWYER [a], Elisabeth A. PIMENTEL [a], Jennifer B. FELLOWS [a], Ryan L. SCOFIELD [a], Vincent L. ACKERMAN [a], Patrick E. HORNE [a], Alan V. LIU [a], Gerald R. SCHWARTZ [a] and Mark W. SCERBO [b]

[a] *National Capital Area Medical Simulation Center of the Uniformed Services, University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD, 20814*
url: http://simcen.usuhs.mil/
[b] *Old Dominion University, Norfolk, VA*

Abstract. This study examines the effectiveness of two virtual reality simulators when compared with traditional methods of teaching intravenous (IV) cannulation to third year medical students. Thirty-four third year medical students were divided into four groups and then trained to perform an IV cannulation using either CathSim™, Virtual I.V.™, a plastic simulated arm or by practicing IV placement on each other. All subjects watched a five minute training video and completed a cannulation pretest and posttest on the simulated arm. The results showed significant improvement from pretest to posttest in each of the four groups. Students trained on the Virtual I.V.™ showed significantly greater improvement over baseline when compared with the simulated arm group ($p < .026$). Both simulators provided at least equal training to traditional methods of teaching, a finding with implications for future training of this procedure to novices.

1. Background

Health care providers must be taught basic clinical skills prior to performing these skills on patients. One of the most basic (and critical) skills taught to healthcare providers is obtaining venous access via intravenous (IV) cannulation. This type of clinical skills education is typically not part of a standardized curriculum, relying mostly on the initiative of faculty to teach medical students these basic skills [1]. Traditionally, teaching methods for IV cannulation have ranged from students being taught using an orange as a model or practicing on a plastic arm, to practicing the procedure on each other and actual patients. These methods may get the job done but they offer inconsistent educational opportunities, a limited variability of case content for the simulated arm, require the willingness of one's fellow student to be repetitively stuck with a needle, require a high faculty-to-student ratio, and are ultimately cost-ineffective [1,2]. Additionally, once IV cannulation skills have been obtained, the opportunities to become proficient and keep these skills

current has become increasingly difficult due to the decreasing opportunities to practice on patients and the reality that many of these basic skills are delegated to ancillary personnel in the clinical setting [2].

Virtual reality (VR) simulators have been an integral part of aerospace and military training for decades and are starting to have increased utilization in the medical community. Simulators offer a variety of potential benefits to medical education and here, specifically, to teaching IV cannulation. Simulators offer a fail-safe environment in which to learn. The student may practice limitless times and is free to fail without the anxiety of causing pain or injury to an actual patient. Both the student's anxiety and the patient's anxiety are removed from the training environment. These devices typically require minimal faculty involvement beyond an initial orientation to operating the product. Thus, the simulator may be more cost effective without even considering the costs associated with the necessary IV supplies and disposal requirements. Additionally, many different levels of health care providers can be trained on a single machine, possibly reducing the cost further. It also may be possible for those training to insert IV catheters to move farther up the learning curve prior to real patient interaction [2]. As simulators become a technology increasingly available to medical educators, it has become important to validate these different systems to ensure tomorrow's health care providers are being properly trained.

There are at least two specific VR systems for teaching IV cannulation available today. CathSim™ (Immersion Medical, Gaithersburg, MD) has been available for 5 years. To date, limited attempts to validate this system have failed to show an advantage over traditional methods of education [2,3]. An additional VR system for teaching IV cannulation, the Virtual I.V.™ (Laerdal Medical, Gatesville, TX) has recently become available and has yet to undergo extensive testing.

The purpose of this study is to examine the effectiveness of these two VR simulators when compared to the traditional modes of using a plastic arm and students practicing on each other for training 3rd year medical students (MS-III) how to perform IV cannulation.

2. Methodology

2.1. Materials

2.1.1. Plastic Simulated Arm

A Laerdal Multi-Venous IV Training Arm was used for the pretest, the posttest and as a training modality. The simulated arm has a layer of plastic skin covering a network of latex veins. In the simulated arm, venipuncture is possible in the antecubital fossa and dorsum of the hand and the accessible veins include the median, basilic and cephalic veins. Artificial blood is connected to the arm. Gravity draws the artificial blood into the venous system of the simulated arm, allowing for flashback when the IV catheter is properly inserted. Additional materials used with both the simulated arm and with the group of students who practiced on each other include latex gloves, tourniquets, alcohol swabs, gauze pads, 20-gauge IV catheter needles, 3-mililiter sterile syringes for flushing the catheter, sterile saline, tape and a biohazard sharps container for proper disposal.

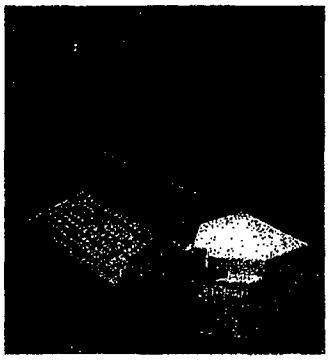


Figure 1. The CathSim™ System showing the graphical user interface.

Figure 2. CathSim™ AccuTouch® Tactile Feedback device allows for tactile (haptic) interaction with the patient on the screen.

2.1.2. *CathSim™*

The CathSim™ system is available from Immersion Medical, Inc. The CathSim™ system (Figure 1) is a microcomputer-based simulator originally developed by HT Medical Systems, Inc., in collaboration with the Department of Nursing, Food and Nutrition of Plattsburgh State University of New York [4,5]. The CathSim™ system provides training on IV catheterization. The physical system consists of an IBM-compatible computer accompanied by an AccuTouch® Tactile Feedback device, a six-degree-of-freedom haptic feedback device, which simulates the catheter needle/hub assembly and a section of the skin for traction. The AccuTouch® Tactile Feedback device (Figure 2) is designed to allow students to experience the tactile responses associated with inserting a needle into the skin and vein. The student also receives audio feedback in the form of patient vocalization that ranges from a crying baby as the IV is inserted to an adult patient saying "ouch!" [3,6].

CathSim™ has a variety of cases for teaching IV catheterization. The patients in each case have different levels of difficulty. For instance, there is an adult male with no complications, and pediatric and geriatric cases with varying complications. A student first selects a patient using the computer mouse. Upon choosing a case, the student must then choose an appropriate site for insertion. The selected insertion site appears on the screen and the student uses a computer's mouse to apply a tourniquet, palpate the vein and cleanse the site by pointing and clicking and dragging objects from a menu of IV catheterization supplies on the screen. Then the student selects the appropriate gauge needle and using the mouse and the AccuTouch® Tactile Feedback device positions the needle and applies the skin traction. Skin traction is simulated by pulling down in the rubber stripping portion of the device with the thumb (Figure 2). Next, the needle is inserted by first fully retracting the simulated catheter and needle of the AccuTouch® Tactile Feedback device and then inserting it as an IV catheter should be inserted. The student is simultaneously looking at the computer monitor to ensure vein access and confirmation of blood backflow. The student then withdraws the needle to complete the procedure. The system automatically ends the simulation once the needle has been withdrawn from the catheter. CathSim™ records many different performance metrics [5,6].

Figure 3. Virtual I.V.™ haptic interface. The lower portion allows for stretching the skin and inserting the needle, and the upper for palpating the vein and applying pressure.

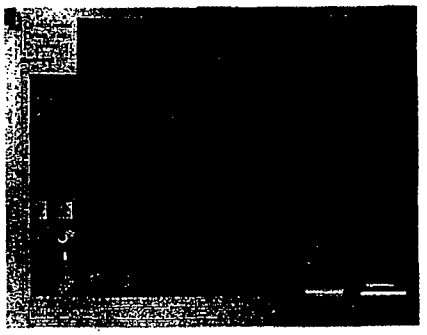


Figure 4. Virtual I.V.™. This screen capture depicts the IV catheter in the vein with bleeding that has resulted from failure to remove the tourniquet and apply pressure.

2.1.3. Virtual I.V.

The Virtual I.V.™ system is available from Laerdal Medical. It is a VR simulator designed to train students on IV cannulation. The physical system is an IBM-compatible computer accompanied by a haptic interface that simulates the catheter needle/hub assembly and two sections of skin to allow for palpation and application of pressure (to stop bleeding) as well as skin traction. The haptic interface (Figure 3) is designed to support palpation, skin stretch and needle insertion with forces dependent on the scenario. The students also receive feedback in the form of bleeding, bruising, swelling, as well as other patho-physiological reactions [7].

The Virtual I.V.™ provides greater case depth than the CathSim™ with over 150 distinct case scenarios. The student can choose from four disciplines: nurses, doctors, EMTs and military care providers. Each patient case is customized to be specifically relevant to one of these disciplines. Then within each discipline are cases with increasing levels of difficulty [7].

A student first selects a discipline and then a level of difficulty within the discipline. A brief history of the patient and why he or she needs an IV is presented. The student must then select the appropriate supplies and the appropriate quantities of these supplies from a menu that contains over a dozen virtual supplies to include biohazard containers and the correct needle gauge for the patient case. The student is then presented with the limb in which the IV catheter will be placed. The student must select the appropriate site on the limb and palpate using the uppermost area of skin on the haptic interface. Then the site must be prepared using the supplies he or she gathered earlier. The student "uses" these supplies by using the computer mouse to click and drag supplies to the appropriate locations. The student then must use the haptic interface to retract the skin using the lowermost skin area and then insert the catheter (Figure 3). Once inserted and the needle is removed, if pressure is not applied to the upper skin area on the haptic interface the patient will bleed (Figure 4). The Virtual I.V.™ system records and evaluates a student's performance using varied performance metrics [7].

2.2. Participants

The participants in this study were thirty-four 3rd year medical students at the Uniformed Services University (USU) selected on the basis of prior IV experience. The students participated as part of their introduction to the third-year clinical surgical clerkship. They ranged in age from 23 to 39 years old (Average =26). The study was approved as an exempt protocol by the Investigational Review Board at USU.

All participants were novices, meaning they had never started an IV. A few students had significant experience with other types of simulators, e.g. flight, driving. None of the students had previous experience with other medical simulators and all students were regular computer users.

2.3. Procedure

All thirty-four third-year medical students viewed a 5-minute training video as a part of their normal curriculum on IV catheterization. All students were asked to fill out a background questionnaire. The students then completed a cannulation pretest on the simulated arm. Performance on the pretest was assessed using a modified version of a standard instrument used to certify this procedure. The instrument is based on a task analysis of the procedure and is scored from 0 to a maximum of 82 points. The students were then randomly divided into four groups. The first group (Each other or EO) consisted of 13 students who practiced IV cannulation on each other for a period of one hour with one faculty member helping and instructing each pair of students. The second group (Virtual I.V.™ or V) of 6 students had up to one hour to practice on the Virtual I.V.™. The third group (CathSim™ or CS) of 7 students had up to one hour to practice on the CathSim™. The final group (Simulated Arm or A) of 8 students practiced for an hour on the IV arm. Upon completion of training, all 34 students performed a post-test on the IV arm within 72 hours of completion of training. Their performance was assessed with the same instrument used for the pretest. The differences in performance between the groups were evaluated using ANOVA and paired t-tests with α set at $p < 0.05$.

3. Results

The results of the pre- and posttest and the change in score (delta) from pre- to posttest is shown in Table 2. There was significant improvement from pretest to posttest overall ($p < .00001$) and in each of the four groups (Table 1).

Comparison of each group with all others revealed no significant difference in pre- or posttest scores between the groups (by ANOVA). However, comparison of the delta or improvement from pre- to posttest revealed that the students practicing on the Virtual I.V.™ had significantly greater improvement than the traditional IV arm group ($p < .026$) and though not significant trending towards it for the EO group ($p < .079$) and the CS group ($p < .058$) (Table 2).

There were no significant differences in performance found based on the gender or age of the participating students. Additionally, there was no correlation between the performance of the four groups based on any previous exposure to non-medical simulators or computer usage.

Table 1. Mean Assessment Scores on pre- and posttest overall (all) and for each of the groups. EO=each other; V= Virtual I.V.™; CS= CathSim™; A=Simulated Arm. All Values are listed as the mean + or – one standard deviation.

Group	n	Pretest %	Posttest %	p value	Delta
EO	13	60.5 (9.7)	73.8 (6.6)	< .0003	13.2(11)
V	6	55.0(12.1)	75.7(3.7)	< .0003	20.7(7.8)
CS	7	66.6(11.8)	78.4(5.2)	< .02	11.9(12)
A	8	65.3 (6.0)	74.9(8.1)	< .009	9.6(10.3)
All	35	63.7(10.1)	76.4(4.6)	< .00001	

Table 2. Comparison of the delta or improvement from pre- to posttest among groups. Values reflect p value with significance at $p < .05$.

	EO	V	CS	A
EO		p < .079	p < .41	p < .24
V			p < .058	p < .026
CS				p < .36

4. Discussion/Conclusions

A review of the limited previous literature of Virtual Reality (VR) simulators has shown that simulators in general are inferior to the traditional methods (the plastic simulated arm) of teaching IV cannulation [1,2,8]. Additionally, no study was found to have compared VR simulators with the method of students practicing on each other to gain experience in IV placement. The results of this study indicate that all four methods of teaching IV cannulation were effective in teaching 3rd year medical students by virtue of improvement over the baseline assessment. Of note is the significantly greater improvement over baseline the students using the Virtual I.V.™ had when compared with those students who learned by practicing on the plastic simulated arm. This finding needs to be confirmed with additional subject accrual. However, this is of particular importance considering that previous studies comparing the CathSim™ to a plastic simulated arm found the students were better trained using the simulated arm. Additionally, one might expect that the students practicing on the simulated arm to have an advantage over the Virtual I.V.™ group as the assessments were performed on the same type of plastic simulated arms. Despite this possible advantage, the Virtual I.V.™ trained students had significantly greater improvement over baseline than those students trained on the plastic simulated arms.

When comparing the other training modes used to teach students to perform IV cannulation, there was no statistically significant difference found. This in and of itself is an important finding as it indicates that both simulators provided at least *equal* training to the more costly and faculty dependent traditional modes of teaching IV cannulation. This means that environments where faculty involvement is constrained have an additional option when considering how to best to maximize their educational resources.

While the results of this study suggest that VR simulators are useful tools in training health care providers, the underlying question is whether the significant improvement all students had over the baseline assessment regardless of training method will translate into improved performance on actual patients. This remains to be seen and is the focus

of future study. Additional areas of focus for future study should include measures of skill degradation over time with interval testing in the four training groups. One of the limitations of this study is that the students were not trained to proficiency. This was due to the limited time these students had available before beginning the third year Surgery Clerkship; however, this deserves emphasis in future studies. Another limitation is the small number of participants. Clearly further study and subject accrual is required to delineate the role of these potentially valuable tools for training.

References

[1] Prystowsky JB, Regehr G, Rogers DA, Loan JP, Hiemenz LL, Smith KM. A virtual reality module for intravenous catheter placement. The American Journal of Surgery 1999; 177:171-5.

[2] Engum SA, Jeffries P, Fisher, L. Intravenous catheter training system: Computer-based education versus traditional learning methods. The American Journal of Surgery 2003; 186:67-74.

[3] Scerbo MW, Bliss JP, Schmidt EA, Thompson SN, Cox TD, Polan HJ. A comparison of the CathSim™ system and simulated limbs for teaching intravenous cannulation. In J.D. Westwood et al. (Eds.), *Medicine meets virtual reality* 2004; 340-6. Amsterdam: IOS Press.

[4] Barker VL. CathSim™. In J.D. Westwood et al. (Eds.), *Medicine meets virtual reality* 1999; 36-37. Amsterdam: IOS Press.

[5] Ursino M, Tasto JL, Nguyen BH, Cunningham R, Merril GL. CathSim™: An intravascular catheterization simulator on a PC. *In J.D. Westwood et al. (Eds.), Medicine meets virtual reality* 1999; 360-6. Amsterdam: IOS Press.

[6] http://www.immersion.com/medical/products/vascular_access/.

[7] http://www.laerdal.com/document.asp?subnodeid=6473945.

[8] Chang KK, Chung JW, Wong TK. Learning intravenous cannulation: A comparison of the conventional method and the CathSim™ intravenous training system. Journal of Clinical Nursing 2002; 11: 73-78.

Appendix B: Publications Supporting Task 2 – Technology Development

1. Frederic D. McKenzie, Hector M. Garcia, Reynel J. Castelino, Thomas W. Hubbard, John A. Ullian, Gayle A. Gliva, "Augmented Standardized Patients Now Virtually a Reality," *Third IEEE and ACM International Symposium on Mixed and Augmented Reality (*ISMAR'04), November 2-5, 2004, Arlington, VA, USA, pp 270-271.

2. Ross R. Vickers, Jr., "Patterns of Disease in the U.S. Military: Looking Back, Thinking Ahead," Unpublished Manuscript, Warfighter Performance Department, Naval Health Research Center, San Diego, CA, 2004.

B.1 Frederic D. McKenzie, Hector M. Garcia, Reynel J. Castelino, Thomas W. Hubbard, John A. Ullian, Gayle A. Gliva, "Augmented Standardized Patients Now Virtually a Reality," *Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04),* November 2-5, 2004, Arlington, VA, USA, pp 270-271.

Augmented Standardized Patients Now Virtually a Reality

Frederic D. McKenzie, Hector M. Garcia, Reynel J. Castelino
Virginia Modeling Analysis and Simulation Center (VMASC)
Old Dominion University
fmckenzi@ece.odu.edu, rcast008@odu.edu, hgarcia@odu.edu

Thomas W. Hubbard, John A. Ullian, Gayle A. Gliva
Theresa A. Thomas Professional Skills Teaching & Assessment Center
Eastern Virginia Medical School
HubbarTW@evms.edu, UllianJA@evms.edu, GlivaGA@evms.edu

Abstract

Standardized patients (SPs), individuals who realistically portray patients, are widely used in medical education to teach and assess communication skills, eliciting a history, performing a physical exam, and other important clinical skills. One limitation is that each SP can only portray a limited set of physical symptoms. Finding SPs with the abnormalities students need to encounter is typically not feasible. This project augments the SP by permitting the learner to hear abnormal heart and lung sounds in a normal SP.

1. Introduction

To become clinically competent physicians, medical students must develop knowledge and skills in many areas of both the art and science of medicine. Three areas are emphasized in medical students' early clinical training: doctor-patient communication, eliciting the history, and performing the physical exam. Standardized patients (SPs), individuals trained to realistically portray patients, are commonly used to teach and assess medical students in those three areas. Working with them provides students the opportunity to learn doctor-patient communication, the history, the physical exam, and other clinical skills in a safe setting. SPs also provide a way to reliably test students' clinical skills in a realistic setting, interacting with a person. The range of clinical problems an SP can portray, however, is limited. They are typically healthy individuals with few or no abnormal physical findings. While some can be trained to simulate physical abnormalities (e.g., breathing through one lung, voluntarily increasing blood pressure, etc.), there are many abnormalities they cannot simulate.

One way to supplement what students learn from SPs is for the students to separately learn from and practice on simulators. A variety of mechanical or computer-based simulators are now used in medical education, including software for testing clinical reasoning and diagnostic skills, computer simulations of physiological processes, and physical models for practicing selected procedural skills. A key limitation is that their users are isolated from interacting with a live person (the patient or SP) while using the simulators. Augmenting SPs with the ability to simulate abnormal physical findings would expand the opportunities for students to learn more clinical skills in a realistic setting with a real person (SP) while practicing their doctor-patient communication skills.

The current phase of this research involves simulating abnormal heart or lung sounds in an SP, thus expanding the breadth of sounds that can be heard in an SP. A learner will listen to an SP's heart and lungs through a modified stethoscope and hear pre-recorded sounds rather than the SP's. With a real or standardized patient, the learner is limited to hearing only the sounds of that single person. Learning a variety of sounds has traditionally required examining many patients over time, often without direct supervision and feedback. Commercially available recordings of heart and lung sounds exist, but using them ignores the process of listening for the sounds (e.g., correct placement of the stethoscope) and excludes simultaneous interactions with the patient. Augmenting SPs with the capability of portraying patients with an increased range of abnormalities will make the use of SPs an even more valuable teaching tool.

2. Methods

Augmented reality integrates text, graphics, sounds, etc. with the natural environment in order to enhance or inform the experience. We have begun by constructing a functional prototype that augments the reality of the SP with virtual sounds for his/her heart and lung. The intent is to impart abnormal pathology that could not be faked or reproduced by the SP alone. This would greatly enhance the repertoire of lessons that could be hands-on trained by the SP and not for their first time on actual patients.

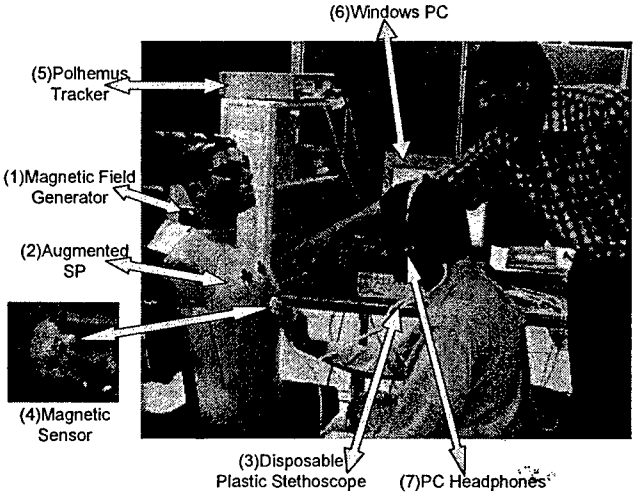


Figure 1. Prototype System

Our simulation involved the use of a mannequin, labeled in Figure 1 as the (2) Augmented SP, fitted with an electromagnetic field generator (1) and a movable sensor connected to the stethoscope head (4). We used the 3SPACE FASTRAK from POLHEMUS as the tracker and the trackd API from VRCO as the PC interface to obtain real time 6 DoF updates of position (X, Y, and Z Cartesian coordinates) and orientation (azimuth, elevation, and roll) [1]. Essentially, the system operates as follows. The medical student would use the disposable stethoscope (3) to place against the Augmented SP and perform auscultation. As the student places the stethoscope head, its position is tracked via the Polhemus FASTRAK (5) using the movement of the attached sensor. When the system software running on the PC (6) detects that the sensor / stethoscope head is placed within an appropriate location, the software triggers the corresponding sound file which plays into the headphones (7) that the student is wearing. Everything occurs in real time and the sounds trigger according to the "hot zones" marked by black crosses on the mannequin.

3. Conclusions

A useable augmented standardized patient system is virtually a reality. We have demonstrated its usefulness by virtue of a proof of concept prototype. A completely virtual SP has been tried before by Hubal et al [2]. They utilize natural language processing, and virtual patients that exhibit emotion in a realistic context to provide completely automatic yet unscripted training sessions. The motivation of the researchers was in replacing the SP because of concerns such as the cost in paying and training actors as well as quality control and reproducibility of a session. These concerns are genuine; however, human-computer interaction brings a different set of psychological concerns than does the human-human interaction of a doctor-patient examination. A significant level of immersion is needed to overcome the human-computer interaction aspects so that there is appreciable transfer of training with regards to patient interaction and diagnosis. This level of immersion and interactivity has not been reached and may not be achievable in a totally virtual form with today's technology. Although Hubal et al provide a useful tool for outlining the steps for patient interaction and diagnosis, our augmented SP work seeks to provide a completely realistic experience by greatly enhancing a currently accepted medium of instruction.

Our project was successful in its attempt to allow listening to pre-recorded heart and lung sounds when the head of the modified stethoscope was placed at any of four locations on a mannequin torso. There is nothing special in the use of the mannequin. A real human (SP) could have been augmented with our system just as easily if she had the inclination to stay still for long periods of time during our testing. The initial proof-of-concept evaluation of the system was performed by an EVMS doctor experienced in SPs and the training of auscultation. Even with the limited functionality of this first prototype, it was evident that the concept works and with a bit of tracking refinement and expanded fields of interest with abnormalities, the system could become a useful and integral part of auscultation education.

4. Acknowledgements

This study was a collaborative project between the Virginia Modeling, Analysis and Simulation Center (VMASC) at Old Dominion University and the Eastern Virginia Medical School. Funding for this study was provided in part by the Naval Health Research Center through NAVAIR Orlando TSD under contract N61339-03-C-0157, entitled "The National Center for Collaboration in Medical Modeling and Simulation". The ideas and opinions presented in this paper represent the views of the authors and do not necessarily represent the views of the Department of Defense.

5. References

[1] www.polhemus.com

[2] Hubal, R.C., Kizakevich, P.N., Merino, K.D., & West, S.L. The Virtual Standardized Patient: Simulated Patient-Practitioner Dialogue for Patient Interview Training. Envisioning Healing: Interactive Technology and the Patient-Practitioner Dialogue. J. Westwood, H. Hoffman, G. Mogel, R. Robb, & D. Stredney (Eds.), IOS Press: Amsterdam, 2000.

B.2. Ross R. Vickers, Jr., "Patterns of Disease in the U.S. Military: Looking Back, Thinking Ahead," Unpublished Manuscript, Warfighter Performance Department, Naval Health Research Center, San Diego, CA, 2004.

Patterns of Disease in the U.S. Military: Looking Back, Thinking Ahead

Ross R. Vickers, Jr.

Warfighter Performance Department
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

e-mail: Vickers@nhrc.navy.mil

Introduction

Military service exposes young healthy individuals to hazardous situations. This paper describes the resulting patterns of illness and disease[1] for United States military personnel. The objective is to identify trends in health care needs.

Disease profiles are central to the general approach to summarizing the available evidence. Profiles describe the health status of a given population. Profiles are defined by a set of rates for each of a number of disease categories. Profiles have two basic components. The term "disease pattern" will be used to refer to the relative importance of different disease categories in the population. This element of the profile is defined by the rank order of disease rates across categories. The term "risk magnitude" will refer to the overall illness burden associated with a given disease. This element of the profile is defined by the rate at which a given disease occurs in the population.

This paper summarizes the literature in terms of profile similarity. The pattern and risk elements of similarity are considered separately to broaden the coverage of the literature. In some cases, risks cannot be compared because studies do not report disease rates. Instead, disease patterns are described in terms of the number or proportion of cases in different categories. This approach often is necessary because the population denominators needed to convert frequencies to rates are unknown. Thus, the information reported in these studies defines a disease pattern, as the term is used here, but does not define risk magnitude. The pattern can be compared to patterns in other populations even though the rates are missing. This comparison is possible because the rank order for different categories is the same whether incidence is described by frequencies, proportions, percentages, or rates. Pattern comparisons cover more of the available evidence than would be the case if comparison required disease rates. Disease patterns therefore are the primary basis for summarizing the evidence.

The evidence summarized in the remainder of this report points to 2 major conclusions. The first conclusion is that two distinct disease patterns can be identified. One pattern describes peacetime hospitalization. The other pattern describes wartime hospitalizations and peacetime outpatient. Thus, a contrast between peacetime and wartime is not only logically relevant to the discussion of military health, but an empirically necessary distinction. Second, when combat status is held constant, specific populations general disease patterns vary somewhat across services and subpopulations within services.

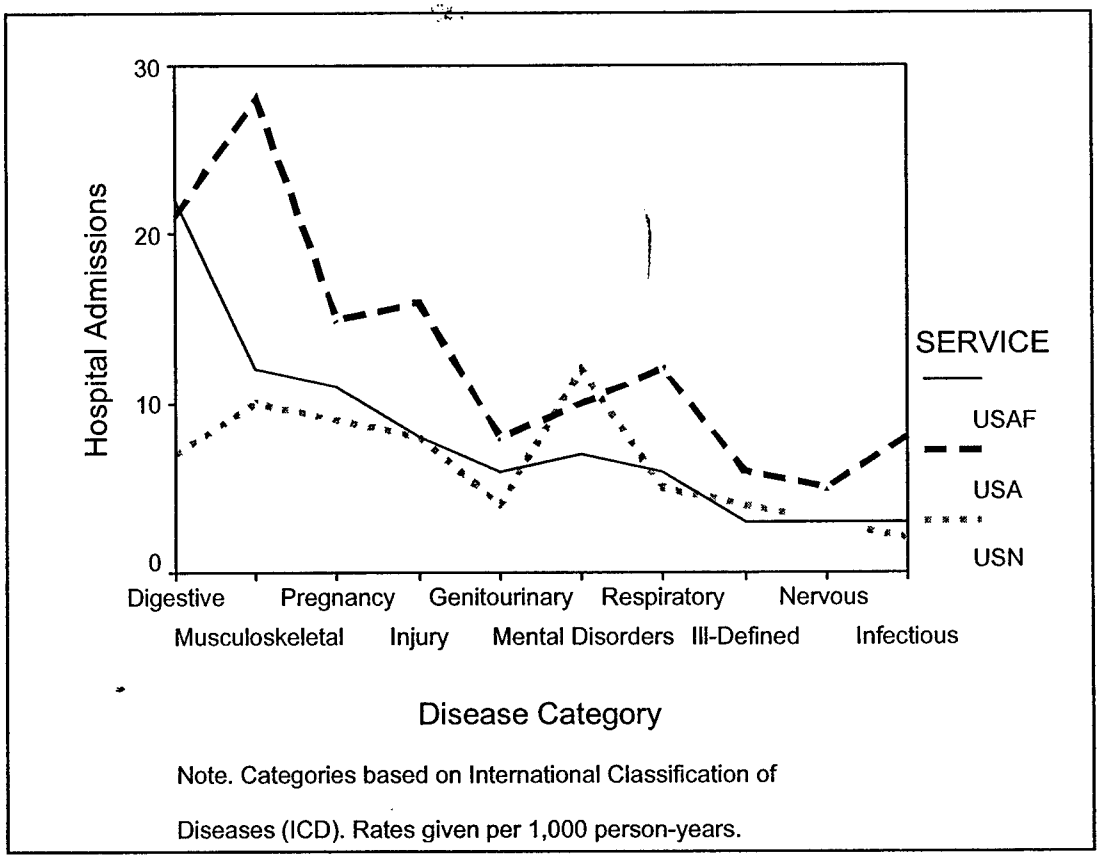
---

[1] The terms "illness" and "disease" are used interchangeably in the remainder of this paper. This usage is in keeping with accepted medical terminology (Spraycar, 1995).

These differences typically do not affect the high and low anchor categories for the disease pattern. Instead, the differences occur among diseases that comprise the middle of the pattern.

An Illustration of the General Approach to Profile Comparison

Disease profiles are central to this review. Figure 1 presents disease profiles for 3 military populations for the year 1992. Each profile represents a set of disease rates that fit the definition of a profile as *the rates of disease observed for different disease categories in a given population during a specific time interval.*



Note. Categories based on International Classification of Diseases (ICD). Rates given per 1,000 person-years.

The profiles shown in Figure 1 provide examples to illustrate 3 basic elements of profile comparison:

- *Pattern Similarity*. The Spearman rank-order correlation (r_s) is computed to quantify pattern similarity.[2] This index is $r_s = 1.00$ for identical profiles. Differences in the rank order of disease categories within profiles produce $r_s < 1.00$.[3]
- *Pattern Discrepancies*. Disease profile differences are described. The disease categories with the largest difference in rankings are identified.
- *Relative Disease Rates*. Disease rates are compared when this information is available. This comparison can be expressed as the ratio of the rate in one population to the rate in the other population.

The profiles in Figure 1 can be compared to illustrate this process. The USAF and USA patterns are similar visually. Rates in both profiles tend to decrease from left to right in the figure. The left-to-right trend is also detectable in the USN pattern, but is less consistent because of the high rate for mental disorders. The similarity coefficients summarize this aspect of the disease profiles. The USAF-USA coefficient, $r_s = .929$ is higher than the USN coefficients (USN-USAF, $r_s = .723$; USN-USA, $r_s = .653$.

Pattern differences can be identified visually in Figure 1. Mental disorders clearly rank higher in the USN profile (1^{st}) than in the other profiles (USAF, 5^{th}; USA, 6^{th}). This difference is the primary reason for the low similarity coefficients for the USN profile. In this case, the comparison is relatively simple and the major basis for the difference is evident. This visual approach would not work well when the number of profiles increased. For example, one analysis reported in the following pages summarizes comparisons among 41 profiles. Another summarizes comparison among 55 profiles. Visual comparison of plots simply is not feasible in such cases.

An examination of risk magnitude completes the profile comparison. Two risk magnitude comparisons illustrate the additional information from this aspect of profile evaluation. First, mental disorder rates were 7 per 1000 person-years for the USAF, 10 per 1,000 person years for the USA, and 12 per 1,000 person years for the USN. The relative rates could be expressed by taking any one of the populations as a reference group. In

---

[2] Rank-order correlations can quantify similarity for health data are reported in different metrics in different studies. The rank of each category will be the same for each metric. The rank-order correlation also is robust to outlier data values.

[3] The computational formula is: $r_s = 1 - 6\sum d_i^2/N*(N^2-1)$, where d_i is the difference in rankings for the i^{th} category.

this case, the pattern differences make comparisons to the USN profile the primary point of interest. Compared to this standard, the relative rates for the USAF and USA were 0.58 and 0.83, respectively.

The second aspect of rate comparisons highlights differences that are not reflected in the pattern comparisons. Musculoskeletal disease consistently was unremarkable in the pattern comparison. This disease category had a high ranking in each profile. However, the relative rates varied widely. The rates for this disease ranged from a high of 28.0 for the USA to a low of 9.7 for the USN with an intermediate value of 12.0 for the USAF. Using the USN as the reference group again, these rates translate to a ratio of 1.24 for the USAF and 2.89 for the USA. These clearly are substantial differences even though musculoskeletal disease ranks highly in all 3 patterns. It is worth noting again that this important element of the comparisons can only be made when rates are reported.

As a final important point, Figure 1 also illustrates another important factor. Profile comparisons must be limited to disease categories that are reported. The data for Figure 1 came from a report that presented the 10 disease categories with the highest rates in each population. This criterion produced the same 10 disease categories for the USAF and USA populations, but the 10th category was different for the USN population. Profiles still could be compared, but this would not have been reasonable if the profiles had only 5 or 6 categories in common.

General Population Hospitalization Profiles

This section compares general population profiles. In most cases, a general population is comprised of all members of 1 branch of service. Less often, the 2 or more branches of service are combined.

General population profiles illustrate the difference between disease patterns for wartime and peacetime. General trends toward lower hospitalization rates make rate comparisons relatively unimportant in these comparisons.

Pattern Stability

Disease forecasting is an important concern in medical planning. Forecasting is simple if profiles are stable. Profiles are stable if both the pattern and rate elements are constant over time. In such a case, yesterday's profile applies to today and the future. Forecasting only requires allowances for chance variation in rates about this fixed core.

Forecasting is more complex if chance is not the only cause of disease profile variation. In this case, other sources of

variation must be identified and quantified to make reasonable forecasts.

Pattern stability is a starting point for assessing profile stability. If the disease pattern is constant over time, the relative importance of different disease categories is constant. Pattern stability is possible if rates either fluctuate in relatively narrow ranges around average values or if all rates show similar patterns of change. For example, rates generally declined between 1980 and 1990.

Gardner, Amoroso, Grayson, Helmkamp, and Jones (1999) provided data that can be used to assess disease pattern stability. These authors reported disease profiles for the U.S. Air Force (USAF), U.S. Army (USA), and U.S. Navy (USN). Profile stability can be examined because separate profiles were reported for each year from 1980-1994 for the USAF, 1981-1994 for the USA, and 1980-1992 for the USN. Profiles consisted of hospitalization rates for the 10 most common disease categories within each service. Diagnoses were classified according to the International Classification of Disease, 9th revision, (ICD-9).[4] Rates were reported for principal disease categories (PDC) within this classification. The central issues regarding these profiles were:

- Are profiles consistent over time within populations?
- Are profiles similar across populations?

Profile similarity coefficients were grouped into 3 categories to answer these questions. The category labels reflect the source of variation that could affect profiles:

Time: The r_s coefficients for disease profiles for different years in a population.

Population: The r_s coefficients for profiles from different populations in the same year.

Population x Time (P x T): The r_s coefficients for pairs of profiles when both the year and population differed.

Table 1 (see following page) presents the average similarity coefficients for these 3 comparison categories.

Profiles appeared to be population specific. First, profiles were moderately stable over time within the 3 populations (r_s = .861). Second, profiles were less similar when compared across populations. The difference was evident even when the comparisons were within the same year (r_s = .703).

[4] The notation "ICD-k" is used in this report to describe the k^{th} revision of the International Classification of Diseases.

Table 1. Disease Pattern Similarity

	Time[a]	Population[b]	P x T[c]	Total
Full Profile	.861	.703	.668	.732
Without Pregnancy	.904	.707	.684	.755

[a]Average r_s within the population across time.
[b]Average r_s within year across populations.
[c]Average r_s when both population and year differed.

One disease category, Pregnancy, had a substantial effect on profile stability. When this category was dropped from the profile (Line 2 of Table 1), average stability increased to r_s = .904. Comparisons across populations were largely unaffected. The term "population-specific" applies to the profiles because each population had a reliable profile that was at least moderately distinct from the profiles in other populations.

U.S. Marine Corps (USMC) Profile. The prior comparisons did not include USMC personnel because Gardner et al. (1999) did not report ICD profiles for this population. However, they did report USMC profiles based on Major Disease Category (MDC), a system used in Navy medicine.[5] Gardner et al. (1999) also reported MDC profiles for the USN population. These two sets of profiles made it possible to perform a USN-USMC comparison that paralleled the prior USAF/USA/USN comparison.

Table 2. Disease Pattern Similarity for USN and USMC

	Stability[a]	Population[b]	Hetero-genous[c]	Total
Males[d]	.917	.850	.827	.869
Females[d]	.945	.951	.943	.944

[a]Average r_s within the population across time.
[b]Average r_s within year across populations.
[c]Average r_s when both population and year differed.
[d]Based on Major Diagnostic Categories for 1989 - 1994 (Gardner, et al., 1999, p. 5-60 and p. 5-73).

[5] MDC-ICD differences include: MDC assigns most injuries to categories based on the affected body component. Sprains, strains, and broken bones would be musculoskeletal disorders. These events would be injuries in the ICD system. MDC separates alcohol and drug abuse from other mental/behavioral disorders; ICD does not. MDC has separate categories for male and female reproductive system disorders; ICD treats both as genitourinary diseases.

The MDC profiles for males and females were analyzed separately because these profiles had only 8 categories in common. These 8 categories represent approximately one-third of the full set of categories in the MDC disease classification system. This representation was too small to be confident that the profiles were adequate indicators of the disease profile. The 10-category profiles at least were comparable to the earlier ICD analysis.

The decision to separate men and women had the serendipitous effect of identifying a gender difference:

- For males, similarity was greater when comparisons were made across time within a population (r_s = .917) than when they were made across populations with year held constant (r_s = .850).
- For females, similarity was essentially the same for comparisons within populations across time (r_s = .945) or across populations within a given year (r_s = .951).

The male pattern was consistent with stable, population-specific profiles. The female pattern was consistent with a stable profile that represented both populations.

Summary of Pattern Stability Findings. Disease patterns are stable over time. The average stability was $r_s > .900$ for men and for mixed populations with the Pregnancy category dropped from the analysis. The average stability was $r_s > .940$ for women even with Pregnancy include. The effect of pregnancy on the overall pattern probably reflects increasing recruitment of women during the 1980s and changes in policies regarding pregnancy. At one time, pregnancy would have been reason for discharge from the service.

Disease patterns were population-specific for men. This characterization is based on the observation that similarity coefficients representing stability within populations were higher on average than coefficients characterizing similarity across populations. The virtual equivalence of the similarity coefficients in the within- and between-population comparisons for women suggests that a single profile applies to the female populations in different branches of the service.

Temporal Trends

Average stability coefficients are misleading. Average within-population similarity was high in the prior analyses because the typical interval between profiles was short. Similarity decreases as the time interval between the profiles

increases (Figure 2). The regression slopes[6] in Figure 2 show roughly equal rates of change for USAF and USA profiles. USN change was somewhat more rapid:

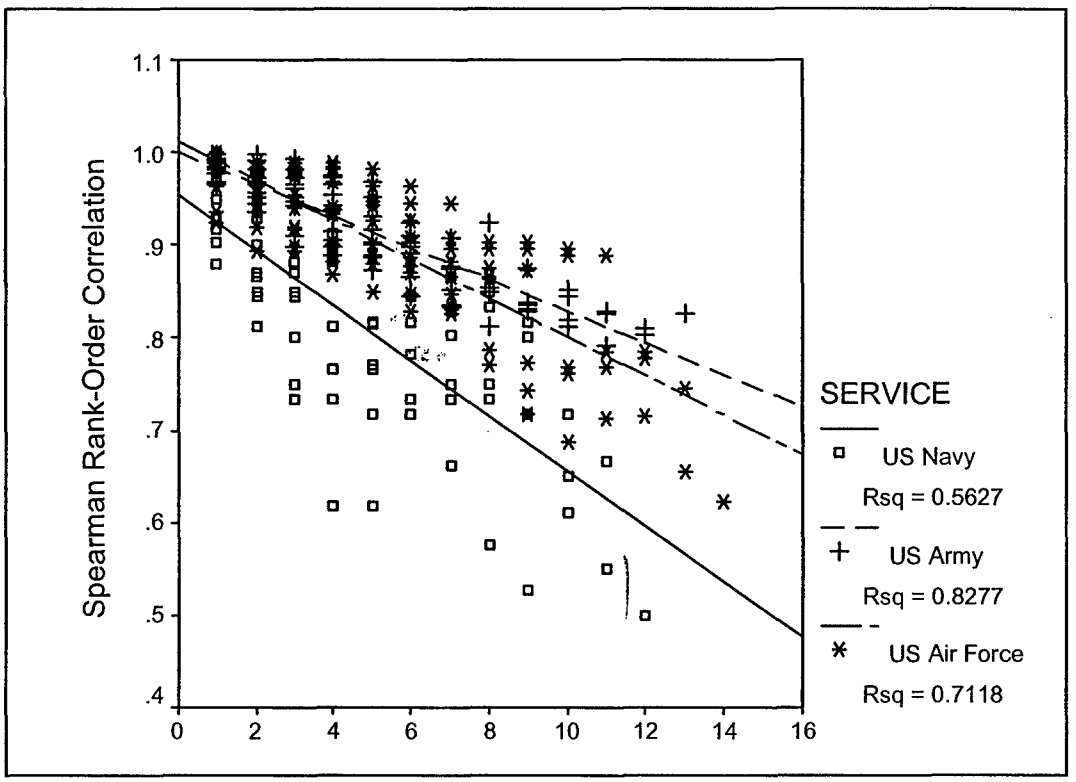


Figure 2. Stability Coefficients as a Function of Time Interval

USAF: r_s = 1.012 - (.021*interval), adjusted R^2 = .709;
USA: r_s = 1.001 - (.017*interval), adjusted R^2 = .826;
USN: r_s = .955 - (.030*interval), adjusted R^2 = .557.[7]

[6] A linear model was adopted after visual inspection of LOESS plots of the data and fitting curvilinear (i.e., quadratic, cubic) and nonlinear (i.e., exponential decay) models to the data.

[7] Significance tests are not reported because they are of limited value in these analyses. Associations that explain ≥26% of the variance would be statistically significant in these analyses. This effect is moderate in size by accepted standards (Cohen, 1988), but it would be too weak to be of much interest for the present purposes. Also, the significance tests would be biased toward leniency. The data are time series, so correlated errors could occur. Correlated errors lead to inflated significance tests (Ostrom, [ref]). In the final analysis, the claim that the trends reflect real phenomena is based on logic and the fact that the empirical evidence was clear, strong, and replicated.

Table 3. Interservice Trend Comparisons

	R^2 Full[a]	ΔR^2 Interaction[b]	b_0[c]	b_1[d]	% Rate of Change[e]
Pregnancy	.945	.043			
Air Force			9.87	.18	1.74
Army			8.68	.47	5.19
Navy			3.22	.50	13.48
Respiratory Disorders	.925	.047			
Air Force			11.85	-.49	-4.31
Army			26.22	-1.05	-4.17
Navy			7.61	-.27	-3.66
Injury	.951	.031			
Air Force			19.18	-.84	-4.58
Army			27.10	-.75	-2.85
Navy			26.76	-1.32	-5.19
Ill-Defined Conditions	.770	.531			
Air Force			7.38	-.31	-4.45
Army			4.15	.13	2.93
Navy			6.54	-.21	-3.39
Infectious & Parasitic	.685	.227			
Air Force			13.12	-.76	-6.12
Army			9.48	.00	.02
Navy			5.75	-.07	-1.24
Musculoskeletal System	.958	.173			
Air Force			12.60	-.03	-.20
Army			12.67	1.11	8.02
Navy			15.12	-.47	-3.21
Mental Disorders	.902	.126			
Air Force			12.77	-.45	-3.62
Army			11.64	.01	.08
Navy			20.02	-.69	-3.55
Digestive System	.974	.085			
Air Force			31.30	-.57	-1.86
Army			11.21	.62	5.26
Navy			11.90	-.37	-3.21
Genitourinary System	.901	.070			
Air Force			9.19	-.23	-2.59
Army			7.76	.02	.25
Navy			5.49	-.13	-2.44

Note. Nervous system disorders not included because data were not available for the USN population. Rates of change for this disorder were close to zero (Army, .00; Air Force, -.05)

[a]Model with interaction

[b]Loss of variance explained from removing interaction

[c]Intercept with year rescaled by subtracting 1980 from the actual year. The intercept, therefore, corresponds to the estimated rate for 1980.

[d]Slope

[e]Slope as percentage of estimated 1980 rate

Decreasing similarity in the disease profiles over time implies differential rates of change in the hospitalization rates for different diseases.[8] Examination of trends in disease rates provided insight into this instability.[9] The most important findings were:

- Trend differences across services accounted for substantial variance (median = 8.5%, range = 3.1% to 53.1%).[10]
- The most common trend pattern was declining USAF and USN rates with stable or rising USA rates (cf., Table 3).
- Injury and Respiratory rates decreased over time in all 3 services.
- Only the Pregnancy rate increased over time in all 3 services.

Summary. Population profiles are highly stable only over short periods of time. Over longer periods, clear trends in the rates modify the profiles. Rising pregnancy rates were the most single factor in changing profiles during the 1980s.

Trend Extrapolation

Future disease profiles could be estimated by trend extrapolation. This approach would produce questionable results in the present case. For example, if the equations in Table 3 equations were used to predict hospitalization rates for 2004 produced, several predicted rates would be less than zero.

The U.S. Army Medical Surveillance Activity (AMSA) has published a series of monthly reports produced by the U.S. Army Center for Health Promotion and Preventive Medicine. These reports document an ongoing program that monitors health trends in the U.S. military. Since 1995, AMSA has published summaries of trends related to specific military health issues. Those publications include annual hospitalization rates covering the

---

[8] If rates of change are identical, trend lines will run parallel. The profile only changes when trend lines cross. When lines cross, the disease category that previously had a lower rate now has a higher rate than the other category.

[9] The analysis was a multivariate analysis of covariance (MANCOVA). Service (USAF, USA, USN) was the group classification variable. Year was the covariate. The design was time, group, and group by time. The repeated measures aspect of the data was disregarded.

[10] The initial service comparisons suggested that different trends for the Army were the basis for service differences. Follow-up analyses compared just the USAF and USN trends. Service differences accounted for ≥1.0% of the variance in 7 of 8 comparisons (median = 2.5%). Large increases in variance explained for Infectious and Parasitic (24.6%) and Musculoskeletal System (27.0%) resulted in an average of 7.6%. Overall, USAF and USN trends clearly differed.

period from 1992 through 2003. These rates may be affected by failure to report some hospitalizations, the use of civilian health care, and other factors. Nevertheless, the rates provide the opportunity to directly extend the trend analysis for one service.

The AMSA profiles also provided a further evaluation of pattern stability and specificity. From 1994 through 1999, patterns were highly stable for men (r_s = .963) and women (r_s = .943). These stable profiles differed substantially whether comparisons were within the same year (r_s = .774) or represented different years for men and women (r_s = .748). The results clearly were consistent with population specific patterns.

For the combined male and female USA population, average stability was high (r_s = .956). Similarity decreased at a rate of ~.018 per year (1 yr, r_s = .980; 2 yr, r_s = .961; 3 yr, r_s =.931; 4 yr, r_s = .916, 5 yr, r_s = .908). Prior analysis indicated a rate of change of .017 per year, so the observed 5-yr similarity was slightly lower than the predicted 5-year similarity (.916). This difference is trivial, so the extension up to 1999 can be regarded as an extension of the earlier trends.

Extending the time period provided a clear illustration of the difficulties with extrapolation. Over the longer period, trends generally were curvilinear. The pattern of increasing rates noted previously gave way to a period of consistent decrease. When this broad trend was modeled mathematically as a quadratic function, the curvature in the function accounted for an additional by 31.8% of the variance in rates on average. In fact, the curvature accounted for >33% of the variation for all but 2 disease categories (Respiratory, 0%; Injury 4.9%).

Table 4. Trends in USA Hospitalization Rates: 1980 - 1999

	Model:		Incremental
	Linear	Quadratic	Variance[a]
Infectious	.440	.829	.389
Mental	.234	.649	.415
Nervous	.508	.843	.335
Respiratory	.907	.907	.000
Digestive	.079	.747	.668
Genitourinary	.468	.838	.370
Pregnancy	.004	.622	.618
Ill-Defined	.113	.547	.334
Injury	.925	.974	.049

[a]Increase in variance explained by adding the quadratic term to the model (i.e., Quadratic - Linear).

These analyses illustrate the obvious point that trends can change. Extrapolating from one time period to another can result

in serious errors. Even the longer time frame of these extended analyses does not provide a definitive pattern. At a minimum, this model shares a critical deficiency with the linear models summarized in Table 3. The typical quadratic equation had a positive linear trend and a negative quadratic trend. This combination will predict negative rates at some point in the future. If the time period were extended further, the ultimate pattern might prove to be cyclical, perhaps with decreasing amplitude.

These trend analyses are a reminder that hospitalization rates are the product of a complex set of causal influences. These influences include geopolitical factors, technological advances in weaponry, human factors design of equipment, mutations of viruses, advances in medicine, viral mutations, climatic trends, organizational policies related to personnel readiness, and so forth. No doubt other factors could be added to this grab bag of possible influences on hospitalization rates. However, the key point is that it is unlikely that all of the factors that influence hospitalization rates will show systematic trends in the same direction. Orderly cumulative trends therefore seem unlikely. Extrapolation may remain the best way to forecast the future, but forecasters should be aware that the predictions may have larger errors than suggested by superficial examination of trends in short term data series.

Summary. Trends can be identified in existing data, but these trends are a dubious basis for forecasting future hospitalization rates. Factors that influence disease rates must be identified and their effects modeled to produce accurate forecasts for specific situations. Current models approach this problem by modeling rates as a function of geographic region and combat intensity (Blood & Gauker, 1993).

Hospitalization Profiles

Disease patterns based on hospitalization rates tend to be population specific and change over time within populations. The disease profiles associated with those patterns are examined here to assess the bases for specificity and change.

Table 5 presents peacetime hospitalization profiles based n ICD classifications. The profiles, which are presented for 3 services, are based on 2 5-year periods. The periods bracket the total interval for which Gardner et al. (1999) reported data. Averaging the rates over 5 years decreases the risk that chance factors will affect the apparent profile changes. The use of an average is supported by the fact that patterns change slowly. Examination of the profiles directs attention a number of important points:

Table 5. Peacetime Reference Profiles

	Early	Late	Rate Ratio
USAF			
Infectious	11.00	3.40	0.31
Mental	11.60	7.20	0.62
Nervous	3.40	3.00	0.88
Respiratory	11.00	6.20	0.56
Digestive	29.40	23.20	0.79
Genitourinary	8.80	6.40	0.73
Pregnancy	10.80	12.40	1.15
Musculoskeletal	13.00	12.80	0.98
Ill-defined	6.60	3.40	0.52
Injury	17.00	8.40	0.49
USA			
Infectious	8.00	8.40	1.05
Mental	11.00	11.20	1.02
Nervous	5.00	5.00	1.00
Respiratory	22.40	13.00	0.58
Digestive	13.60	19.40	1.43
Genitourinary	7.80	8.00	1.03
Pregnancy	10.40	14.80	1.42
Musculoskeletal	17.00	27.00	1.59
Ill-defined	4.60	6.00	1.30
Injury	23.80	17.00	0.71
USN			
Infectious	4.98	5.94	1.19
Mental	18.34	13.16	0.72
Nervous	. [a]	. [a]	. [a]
Respiratory	6.98	4.76	0.68
Digestive	10.88	8.08	0.74
Genitourinary	5.14	4.16	0.81
Pregnancy	4.74	8.10	1.71
Musculoskeletal	14.28	10.28	0.72
Ill-defined	5.74	4.36	0.76
Injury	23.16	13.48	0.58

Note. The tabled rates are averages computed from rates reported in Gardner et al. (1999, pp. 5-28 (USA), 5-104 (USAF), and 5-140 (USN). "Early" rate is the average of 1980 through 1984 for USN and USAF and 1981 through 1985 for USA. "Late" rate is the average of 1990 through 1994 for USAF and USA and 1988 through 1992 for USN. The rate ratio is Late/Early.

[a]Hospitalization rates for Nervous were not reported for the USN.

- Patterns changed over time. The early-late similarity coefficients were: USAF, r_s = .773; USA, r_s = .830, USN r_s = .746.
- Pregnancy rates were important. Pregnancy was the only disease category that moved up in the rankings for all 3 populations. The initial pregnancy ranking varied (USN, 9th, USAF, 7th, USA, 6th), but the final ranking was the same in each population (4th). Dropping pregnancy from the analysis, stability coefficients were: USAF, r_s = .912; USA, r_s = .850, USN r_s = .857.
- The most common diseases were similar for all 3 populations. Ignoring Pregnancy, Digestive, Injury, Musculoskeletal were 3 of the top 4 categories except in the later profile for USN. Digestive dropped to 5th in that profile.
- The later USA and USAF profiles were very similar (r_s = .906), but the USN profile was distinctive (USAF, r_s = .687; USA, r_s = .758). The rankings of Digestive (lower) and Mental (higher) made the USN pattern distinctive.
- USAF rates dropped in 8 of 10 categories, while USN rates dropped in 7 of 9 categories. USA rates in 7 of 10 categories, but only 4 increases were >1.0 admissions per 1,000 person-years.
- The cumulative rate for the reported categories dropped 29.5% for the USAF (122.6 per person-years to 86.4 per person-years) and 23.0% for the USN (94.2 per person-years to 72.5 per person-years).[11] The cumulative rate increased by 5.0% for the USA (123.6 per person-years to 129.8 per person-years).

The actual frequency of diseases suggests somewhat less dramatic changes.

- Some disease rates changed dramatically over time. These changes are expressed as rate ratios in Table 5. When the differences are expressed as percentage changes and absolute changes, the outstanding differences were:
  - *USAF.* Hospitalization rates dropped by 51% (8.6 per 1,000 person-years) for Injury, by 69% (7.6 per 1,000 person-years) for Infectious, and by 21% (6.2 per 1,000 person-years) for Digestive.
  - *USN.* Hospitalization rates dropped by 42% (9.8 per 1,000 person-years) for Injury, by 28% (5.1 per 1,000

[11] Changes in reporting requirements affect the USN estimates. After 1988, hospitalizations were not reported for deployed troops (Gunderson, [ref]). This change may have reduced rates by 15% (Gunderson, personal communication, 2004). Because this changed applied to 4 of the 5 years contributing to the late profile, the average effect would be ~12%. The true drop in USN rates, therefore, may be closer to 10%.

person-years) Mental, and by 28% (4.0 per 1,000 person-years) for Musculoskeletal.[12]

- *USA*. Hospitalization rates increased by 59% (10.0 per per 1,000 person-years for Musculoskeletal and by 43% (5.8 per 1,000 person-years) for Digestive. Rates decreased by 42% (9.4 per 1,000 person-years) for Respiratory and by 29% (6.8 per 1,000 person-years) for injury.

- Injury was the only category with a rate that declined in all 3 services (USAF, 51%; USN, 42%; USA, 29%).

Table 5 can be recast as comparisons between services (Table 6). The outstanding points in these comparisons included:

- The Respiratory hospitalization rate for the USA consistently was more than twice those of the USAF and USN.
- The high rate of Digestive hospitalizations in the USAF relative to the USN was the only other instance of a ratio that exceeded 2.00 in both periods.
- USN rates were consistently lower than USA and USAF rates during the early 1990s. This difference may be attributed to changes in reporting requirements. From 1990 onward, deployed USN units did not report hospitalizations (Gunderson, [ref]). This change would inflate the ratios because USN rates were in the denominator. Based on changes in the overall hospitalization rate from the year before the change in requirements to the year after, this modification reduced rates by about 15% (Gunderson, personal communication, August, 2004).

Table 6. Hospitalization Rate Comparisons Between Services

	USAF/USA		USAF/USN		USA/USN	
	Early	Late	Early	Late	Early	Late
Infectious	1.38	.40	2.21	.57	1.61	1.41
Mental	1.05	.64	.63	.55	.60	.85
Nervous	.68	.60	.[a]	.[a]	.[a]	.[a]
Respiratory	.49*	.48*	1.58	1.30	3.21*	2.73*
Digestive	2.16	1.20	2.70*	2.87*	1.25	2.40
Genitourinary	1.13	.80	1.71	1.54	1.52	1.92
Pregnancy	1.04	.84	2.28	1.53	2.19	1.83
Musculoskeletal	.76	.47	.91	1.25	1.19	2.63
Ill-Defined	1.43	.57	1.15	.78	.80	1.38
Injury	.71	.49	.73	.62	1.03	1.26

Note. Table entries are rate ratios computed as indicated in the column labels. Thus, the USAF/USA ratios were computed by dividing the relevant USAF rate by the relevant USA rate.
[a]Hospitalization rates for Nervous were not reported for the USN.

[12] The change in reporting requirements would affect these values. Better estimates might be obtained by subtracting 12% from each value, i.e., 30%, 16%, and 16%, respectively.

These reference profiles cover the time period from 1980 through 1994. This period represents the only time interval for which comparable profiles could be defined for all 3 services.

Wartime Hospitalization Profiles

Wartime reference profiles complement the peacetime profiles. Combat activities should change the disease profile. Injury rates should increase. Infectious disease rates will increase if mobilization exposes more personnel to novel pathogens or if the stress of war down regulates immune function.

Table 7. Similarity of ICD Profiles for Major Combat Periods

World War I	1.000	.983	.927	.867
World War II	.985	1.000	.923	.904
Korea	.810	.804	1.000	.845
Vietnam	.871	.903	.799	1.000

Note. Based on rankings given in Table 1 of Hoeffler and Melton (1981, p. 779). Values below the diagonal are for all 18 disease categories; values above the diagonal are for the 17 categories remaining when supplementary classifications are removed.

Hoeffler and Melton (1981) reported disease profiles for the 4 major 20th Century conflicts involving the United States. The profiles consisted of rates for the 17 ICD (8th revision) disease categories. The profiles also included rates for supplementary codes given to admissions for reasons other than a disease or injury (e.g., vaccination, carrier of communicable disease). The rates were based on administrative data reported annually to the Navy Surgeon General for USN and USMC personnel. The rates were based on world-wide hospitalizations and therefore do not reflect pure combat rates. Actual admission rates were not reported. Instead, the rank orders for the disease categories were given. Analysis of these rankings indicated that:

- The disease profile was virtually unchanged from World War I to World War II (r_s = .985).
- Supplementary codes were important in the Korean War. This category ranked 6th in that profile compared to 15th for World War I and II and 13th for Vietnam. The Korean War profile was very similar to the two World War profiles (World War I, r_s = .927; World War II, r_s = .923) when the supplementary category was dropped from the analysis.
- The Vietnam War profile was distinctive. Much of the distinctiveness arose from differences in the rankings for the Nervous (10th vs. 5th in the World Wars and 8th in Korea) and Mental (6th vs. 11th in the other profiles) categories.

The wartime disease profile was largely unchanged from 1917 through 1953. The Vietnam War saw relatively higher rates of mental disorders and relatively lower rates of nervous system diseases. Despite these changes, the most common diseases were the same in all 4 conflicts:

- Respiratory disease ranked 1st from World War I through the Korean War and 2nd for Vietnam.
- Infectious disease ranked 2nd for World War I through the Korean War and 3rd for Vietnam.
- Injury ranked 3rd for World War I through the Korean War and 4th for Vietnam.

The evidence is clear on two points. During wartime, the Respiratory, Infectious, and Injury categories generate the most hospitalizations. This pattern has been stable for some time. Recent changes in the ranking of the Mental and Nervous categories have been the only substantial modifications in the profile during the 20th century. This pattern stability has been maintained despite substantial changes in overall hospitalization rates. Hoeffler and Melton (1981) noted that hospitalization rates dropped from 800 per person-years to 100 per 1,000 person-years between 1900 and 1975.

Table 8. Similarity of Peacetime and Wartime Reference Profiles

Peacetime Profile[b]	Wartime Profile[a]: World War I	World War II	Korean War	Vietnam War
USAF				
Early	.195	.170	.255	.450
Late	-.280	-.347	-.207	-.097
USA				
Early	.261	.273	.406	.479
Late	-.030	-.055	.042	.176
USN				
Early	.006	.055	.127	.552
Late	-.231	-.176	-.097	.304

[a]Profile from Table 1 of Hoeffler and Melton (1981).
[b]Profile from Table 3 of this report (cf., p. 14).

The substantial stability of the wartime disease pattern contrasts with the relatively rapid changes in the peacetime pattern during the 1980s. Further examination underscores the contrast between peacetime and wartime. The disease profiles for the two periods are distinctive (Table 8). Similarity coefficients ranged from r_s = -.347 to r_s = .552 with a median value of r_s = .149.

The differences between the patterns were striking. The disease categories that were most common that changed. During peacetime, the most common reasons for hospitalization fell in the Digestive, Pregnancy, and Musculoskeletal categories. As noted previously, the Injury, Respiratory, and Infectious categories generated the most hospital admissions.

The pattern differences are particularly striking when it is remembered that the wartime patterns represented all USN and USMC personnel. The wartime patterns represent disease in a mixed population. Some were in combat, some were in combat support, and others were stationed outside the combat zone. Because many people in operational areas are not directly involved in combat, the direct effects of combat per se are limited to a small proportion of the total population. The effects of combat must be dramatic for changes in rates in a small proportion of the population to significantly modify the overall population profile. Note that the difference evidently is not confounded with service differences in disease patterns described earlier. The similarity coefficients for the USN peacetime pattern were as low as those for the other profiles.

Potential Influences on Hospitalization Disease Profiles

The next several sections of this report examine factors that might affect disease profiles. These factors conceptually have a hierarchical structure. Geopolitical factors affect the distribution of military units throughout the world and their activities within each region. Military units operate within specific geographic regions in response to operational requirements. Different units serve specific functions as elements of the overall activities within a region. Operational requirements may mean that similar units perform different missions with different frequencies and different durations in different geographical regions. Personnel in different occupational specialties perform different tasks as part of unit functions. The specific mix of tasks and other factors that could affect disease rates could vary from one region to another.

The ideal study to separate the effects of the various factors that influence disease in military populations would examine the combined effects of geopolitical considerations (e.g., peacetime vs. wartime; region of conflict), unit functions (e.g., military service and unit type within service), operational factors (e.g., length of deployment, operational tempo), unit type (e.g., ship vs. shore for the USN), and individual differences on disease. The ideal study would extend the examination of these factors over time to define trends, examine the effects of variations in operational tempo, and so forth.

No available study has the complexity or duration of the ideal study. Typical studies report health statistics for a single service in a single region for a particular year. Studies that examine subsets of the factors are reviewed below.

Geographic Region

Geographic region is one way to define subpopulations within a larger population. Personnel in different regions are exposed to different climates and infectious disease agents. Personnel in different regions also encounter different operational requirements.

Regional effects have been investigated for the USN and the U.S. Marine Corps (USMC). Pugh, White, and Blood (1989) provided regional hospitalization rates for USN. Hermansen, White, Pugh, and Shaw (1990) provided similar information for USMC personnel. The primary regions were the United States, Europe, and Southwest and Southeast Asia. USN personnel within each region were divided into those assigned to shore stations and those assigned to ships. Marine Corps personnel were not assigned to Southwest Asia. Thus, the analysis considered 3 USMC and 6 USN populations.

Annual ICD-9 disease rates were reported for each population. Similarity coefficients were divided into comparisons within a population over time (Population), between populations within a year (Year), and all others (Other). Profile comparisons are based on average within-category similarity.

The similarity coefficients indicated that disease profiles were population-specific (Table 9). The average within-population similarity (r_s = .889) was higher than the average between-population similarity. The similarity of between-population comparisons was the same whether the comparisons were made within a given year (r_s = .844) or across different years (r_s = .845).

Table 9. Disease Pattern Similarity: USN and USMC Regional Populations, 1980-1984

	Population[a]	Year[b]	Other[c]	Total
Regional Populations[d]	.889	.844	.845	.848

[a]Average r_s within the population across time.
[b]Average r_s within year across populations.
[c]Average r_s when both population and year differed.
[d]ICD data from Hermansen et al. (1990) and Pugh et al. (1989) for 1980 - 1984. See text for details.

Table 10. Hospitalization Trend Analysis: USN and USMC Regional Populations, 1980-1984

Disease/Disorder[a]	Full Model Adjusted R^{2b}	Interaction Variance Explained[c]	Variance Explained per df[d]
Infectious	.790	.049	.005
Neoplasm	.776	.017	.002
Endocrine	.667	.000	.000
Blood	.516	.016	.002
Mental	.936	.027	.003
Nervous	.552	.290	.029
Circulatory	.875	.023	.002
Respiratory	.797	.040	.004
Digestive	.917	.058	.006
Genitourinary	.848	.026	.003
Skin	.617	.150	.015
Musculoskeletal	.811	.090	.009
Congenital	.642	.001	.000
Ill-defined Conditions	.852	.036	.004
Injury	.895	.001	.000

[a] Disease categories have been given 1-word labels to simplify the presentation of results. See Appendix A for full ICD category labels. The Pregnancy and Perinatal categories are missing from the table because were Perinatal rates were 0.00, so there was no variance to analyze. Pregnancy rates were 0.00 in all deployed Navy populations. The resulting zero slopes and variances inflated the interaction term.

[b] R^2 for model adjusted for sampling effects.

[c] Difference between adjusted R^2 for the model with (Full model) and without (Reduced model) the interaction.

[d] The variance explained by trend differences was based on 10 df. The average effect is the expected value for pairwise comparisons such as those in previous analyses.

Trends Within Regions

Temporal trends were relatively unimportant (Table 10). Population-by-time analyses (i.e., 11 populations by 5 years for each diagnostic category) showed:

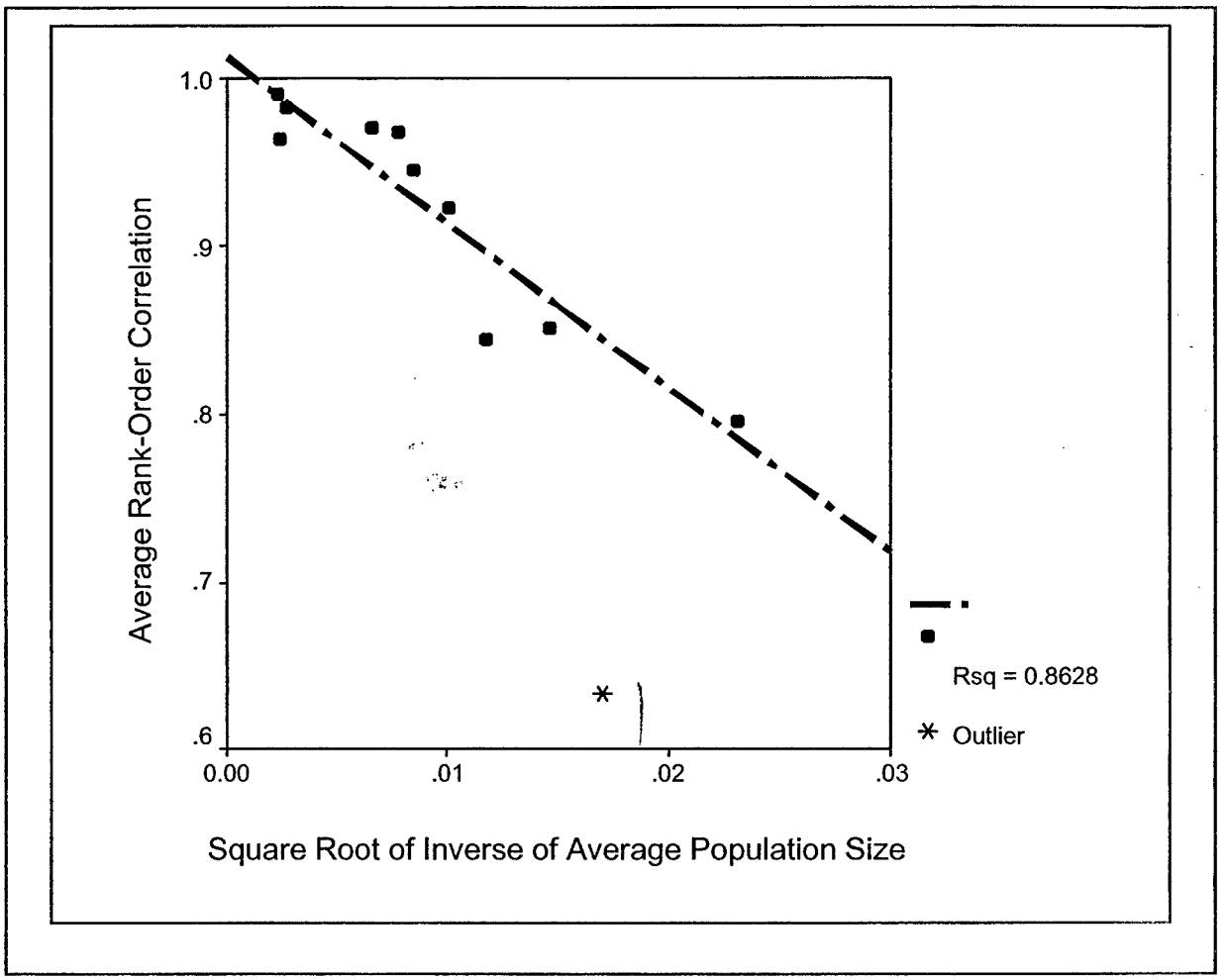
- The models had good overall explanatory power (median adjusted R^2 = .852).
- Trend differences were trivial sources of variance (median variance explained = .027).
- Interactions were trivial. The population by time interaction accounted for less than 9% of the variance in

13 of 15 analyses. Given 10 degrees of freedom in the interaction tests, this amounts to less than 1% of the variance per degree of freedom.

- Two interactions had substantial explanatory power, but:
  - o *Nervous*. The overall interaction for this category resulted from a single atypical rate. The USN rate for Southwest Asia in 1984 (60 hospitalizations per 1,000 person-days) was exceptionally high. When compared to other rates for that population, this atypical rate was 7.5 times the next highest rate. In fact, this atypical rate was 2.4 times higher than any other Nervous system rate in the data. When analysis was limited to the other 8 populations, the interaction was trivial (Variance explained = .023; Variance explained/df = .003).[13]
  - o *Skin*. The interaction for this category combined 2 strong negative slopes (shipboard USN Northeast Asia, b = -1.913; shore USN Southwest Asia, b = -2.468) and 2 strong positive slopes (shore USN Europe, b = 1.465; shore USN personnel Northeast Asia, b = 3.106) with 7 slopes close to zero ($-.111 \leq b \leq .868$). These trends were questionable because none of the exceptional slopes replicated from ship to shore.
- A second set of analyses eliminated the interaction to test for temporal trends that were general across all 11 populations:
  - o Three of 16 diseases produced nominally significant trends representing small (Injury, 2.4%; Circulatory, 1.1%) to moderate (Congenital, 15.2%) proportions of variance explained.
  - o The hospitalization rate decreased over time for all 3 diseases (accidents, -.99 per year; congenital problems, -.22 per year; circulatory disease, -.15 per year).

The central finding in these analyses is that temporal trends accounted for small amounts of variance in the disease rates. The only exception was a disease category that has such low overall rates and limited variation that moderate explanatory power of the temporal trend has little effect on overall disease rates. Thus, temporal trends effectively were absent in these data.

[13] Replacing the questionable rate by 6 per 1,000 person days also eliminated the interaction (variance explained = .021; variance explained/df = .002). Thus, the interaction could the product of a typographical error (i.e., typing .060 in place .006).



Population differences were not as well-defined as in earlier analyses. The similarity of profiles for different years within a population was lower on average. The difference between the within-population and between-population similarities was smaller. Lower similarities within populations might be explained by the smaller sizes of the populations. Smaller populations imply less precise estimates of statistical parameters such as hospitalization rates. Figure 3 illustrates how this factor applies to the present data. However, this explanation would be expected to apply to between-population comparisons, too. Thus, it does not explain why those comparisons are not lower.

Combat Effects

Combat exposure should modify the disease pattern. Combat would be expected to increase injury rates, particularly as the intensity of combat increased (Blood & Gauker, 1993). Combat would also be expected to affect hospitalization rates for mental disease (e.g., post-traumatic stress disorder). It is less clear

what might be expected in other disease categories, but combat also brings changes in living conditions, environmental exposures, and other factors that could affect disease patterns. This section of this report reviews studies that evaluated these possible effects.

Blood and Aboumrad (2001) reported post-deployment hospitalization rates for Vietnam and the Persian Gulf veterans. The Vietnam sample consisted of 11,894 marines who served as members of the 1st Marine Division in Vietnam and were not discharged from the service immediately upon their return. The Persian Gulf sample consisted of 10,878 men who served 1st Marine Division units comparable to those in the Vietnam conflict. Despite these efforts, the Vietnam sample had a higher percentage of infantry (59% vs. 40%) and a lower percentage of service support troops (26% vs. 40%).

Hospital admissions were determined for the 5 years following the end of each war. Diagnoses were coded according to ICD-8 for the Vietnam sample and ICD-9 for the Persian Gulf sample. The Pregnancy and Perinatal categories were omitted, so only 15 disease categories defined the pattern. Population at risk was determined by computing the number of days that each person remained in the service after returning to the United States.

Table 11. Persian Gulf and Vietnam Patterns versus Reference Patterns

	Full Profile[a]:		Top 10 Profile[b]:	
	Vietnam	Persian Gulf	Vietnam	Persian Gulf
World War I	.665	.596	.033	-.200
World War II	.663	.589	.059	-.217
Korea	.769	.700	.293	.067
Vietnam	.808	.732	.427	.150
USAF			.706	.850
USA			.594	.833
USN			.577	.850

Note. Based on data from Blood and Aboumrad (2001). Peacetime profiles were defined by hospitalization rates reported in Gardner et al. (1999, see Table 5, p. 14). Wartime profiles were defined in Table 1 of Hoeffler and Melton (1981).

[a]Computed for 15 categories. Pregnancy, and Perinatal were not reported.
[b]Computed for 9 categories, the 10 most common diseases minus Pregnancy.

The previous analysis of combat profiles indicated that the rate of supplementary conditions was particularly high in

Vietnam. The similarity between this profile and other combat profiles increased substantially when that category was dropped from the analysis (cf., p. 17). The same was true in the Vietnam-Persian Gulf comparison. The profiles were moderately similar (r_s = .861) with supplementary conditions in the analysis and highly similar (r_s = .924) with them removed. Comparison to Hoeffler and Melton's (1981) wartime patterns showed that:

- The profiles were more similar to recent combat profiles than to those early in the 20th century (Table 11).
- For the 10 most common diseases, the profiles were more similar to peacetime profiles than to combat profiles.

Two categories contributed heavily to the similarity between the Persian Gulf profile and peacetime profiles.

- Infectious disease ranked 7th in the Persian Gulf data which is comparable to peacetime rankings (USAF, 6th; USA, 6th; USN, 8th), but lower than typical combat rankings (i.e., 1st to 3rd).
- Musculoskeletal disease ranked 1st, accounting for 21% of all admissions. This category historically has ranked 4th during periods of combat and 3rd in peacetime. This slight difference is amplified in computing In prior conflicts, this disease category typically has been the 4th most frequent reason for hospitalization. During peacetime, musculoskeletal disorders historically have been the 3rd most common reason for hospitalization.

Palinkas and Coben (1988) reported hospitalization rates for USMC personnel during the Vietnam War. Their data, which covered the period from 1965 to 1972, were coded according to the Department of Defense Disease and Injury Codes (DDDIC) for hospitalizations prior to 1970. ICD-8 coding was applied to hospitalizations from 1970 through 1972. Separate rates were computed for personnel assigned to duty in Vietnam and for all other USMC personnel.

The profile for USMC personnel serving in Vietnam was only moderately similar to all other marines (r_s = .744). Combat status affected the Vietnam profiles. The pattern for personnel serving in Vietnam was moderately similar to other combat patterns. The Vietnam pattern was clearly distinct from peacetime patterns. The profile for personnel who not serving in Vietnam was somewhat similar to both sets of reference profiles.

Respiratory disease was a major factor in the Vietnam profiles. The ranking for this category, which was only 9th, contrasted with its usual rank as one of the top 3 categories. With Respiratory removed from the analysis, the disease pattern for Vietnam personnel was clearly more similar to the combat reference profiles (r_s > .718) than to the peacetime reference

Table 12. Effects of Serving in Vietnam on USMC Pattern

Reference Pattern	Full Pattern: Vietnam	Full Pattern: Non-Vietnam	Minus Respiratory: Vietnam	Minus Respiratory: Non-Vietnam
Combat Era				
World War I	.618	.634	.719	.563
World War II	.725	.693	.837	.631
Korean War	.614	.710	.719	.651
Vietnam War	.818	.869	.907	.838
Peacetime				
USAF	.117	.686	.048	.762
USA	.067	.883	.214	.833
USN	.167	.817	.190	.881

Note. Based on data from Palinkas and Coben (1988). Peacetime profiles were defined by hospitalization rates reported in Gardner et al. (1999, see Table 5, p. 14). Wartime profiles were defined in Table 1 of Hoeffler and Melton (1981).

profiles (r_s < .215). Among those who did not serve in Vietnam, removing Respiratory reduced the similarity to combat reference profiles and, on average, increased the similarity to peacetime profiles.

The preceding observations point to respiratory disease as a key element of the disease profile. Table 11 provides a perspective on why this was the case. Overall, combatants were more than 2.5 times more likely to be hospitalized than noncombatants (RR = 2.54). The combat population had a higher hospitalization rate for 13 of 15 disease categories. The combat population rate was more than twice as high in 9 categories. In 5 of those 9 categories, the rate for the combat population was more than 3 times that in noncombatants.

Hospitalizations for Respiratory diseases were not elevated in the combat population. To the contrary, Respiratory disease was 1 of 2 categories for that produced slightly lower hospitalization rates in the combat population. This result contrasted markedly with the results for the other two disease categories that typically differentiate combat disease profiles from noncombat profiles. Injury rates better than doubled (RR = 2.48), but this increase paled when compared to the greater than 6-fold increase in the Infectious category (RR = 6.60).

Table 13. USMC Hospitalization Rates: Vietnam Combat Status

	Vietnam	Non-Vietnam	Relative Risk
Injury	116.9	47.1	2.5
Infectious	93.1	14.1	6.6
Ill-Defined	67.4	12.0	5.6
Skin	47.2	15.6	3.0
Musculoskeletal	43.5	19.6	2.2
Nervous	37.5	10.4	3.6
Mental	35.3	20.0	1.8
Digestive	33.0	15.6	2.1
Respiratory	25.8	35.1	.7
Supplemental	17.9	5.3	3.4
Genitourinary	17.3	7.9	2.2
Circulatory	10.9	7.1	1.5
Neoplasm	6.5	3.6	1.8
Blood	4.7	1.0	4.7
Congenital	3.8	4.0	.95
Endocrine	3.7	2.2	1.7
Total	564.70	220.80	2.56

Note. Based on Palinkas and Coben (1988). Relative rate = Vietnam Rate/Non-Vietnam. Supplemental codes were treated as non-disease hospitalizations.

Occupational Differences

Occupational differences could be examined in some detail for the USN. Hospitalization rates from EPISYS (Jaeger, White, & Show, 1996) for the period from 1980 to 1994 provided the basis for analyzing occupational disease patterns. The patterns were defined by rates for 13 ICD categories (cf., Vickers, 1998).[14] The rates represented 6,408,717 person-years of observation distributed across 82 Navy enlisted occupations. Eight occupations with fewer than 5,000 person years of observation were dropped. This elimination represented 0.2% of the total data (15,069 person years). Another 0.6% of the data (36,203 person years) could not be used in the analyses because the occupational code was missing from the database. With these eliminations, the analysis was based on 74 enlisted occupations represented by 6,357,445 person years of observation.

Comparing 74 patterns would require the examination of 2,701 pairs of patterns. Given this volume of evidence, patterns could be hard to discern and be difficult to report. The following steps were taken to avoid this problem:

[14] The missing categories were Pregnancy, Perinatal, Congenital, and Endocrine.

Table 14. Total Hospitalization Rate by Occupational Category

Category	Disease Rate	Occupations
2	25718.98	SN, FN, AN
4	17454.36	HM, DT
10	8232.71	RP
7	7200.08	SM, OS, BT, IC, HT, GS, CE, CM, SW, UT, AD, AO, AB, AS (k = 14)
5	6962.51	BM, OM, YN, PN, DP, SK, DK, MS, SH, JO, LI, AZ (k = 12)
6	6906.13	QM, TM, GM, MN, IM, RM, PC, EN, MR, DC, BU, PR (k = 12)
9	5747.58	OT, WT, MU, AC, AG, TD, AK, PH (k=8)
8	5645.92	EW, ST, FC, FT, MT, ET, DS, CT, IS, MM, EM, CA, AT, AX, AW, AE, AM (k=17)
1	6904.97	MA, NC, LN
3	6835.82	AF

- Principal components analyses of the hospitalization rates identified 2 large components (λ_1 = 9.53, λ_2 = 2.19) that accounted for 90.1% of the total variance.
- Cluster analyses using the PCA component scores to characterize occupations identified 10 groups with similar profiles.
  - o The initial analysis distributed 9 isolated occupations across 4 clusters; the remaining 64 occupations formed a core cluster.
  - o Analysis restricted to the 64 core occupations produced 6 clusters with one singleton cluster and 5 clusters with between 8 and 17 occupations.
  - o The 10-category classification combined the 4 non-core Stage 1 clusters with the 6 Stage 2 clusters.
- Disease profiles were computed for each occupational cluster using number of person-years of observation as a weighting factor. Analysis of the resulting profiles showed that:
  - o Profiles were moderately similar on average (Kendall's W = .830).
  - o Eliminating a category with 1 occupation (AF), which was represented in the data by only 5,018 person-years of observation, W increased by .091 (W = .921).
  - o Eliminating a second category with 3 occupations (MA, NC, LN), which were represented by 52,541 person years of observation, W increased by .023 (W = .946).

- Profiles from 6 categories that encompassed the original core cluster of 64 occupations were highly similar (W = .965).

The overall picture was clear. Most occupational disease profiles were moderately to highly similar. Cluster differences were defined primarily by differences in the overall rate of hospitalization (Table 12). The highest rates are in apprentice classifications (SN, FN, AN). The incumbents in these occupations have been in the service a relatively short time and have not yet been assigned to specific occupational categories. The next highest rates are found among health care providers, corpsmen (HM) and dental technicians (DT).

The remaining occupations could be characterized as groups with moderate hospitalization rates (38 occupations) or low hospitalization rates (25 occupations).

The preceding generalizations cover 69 of the 74 occupations analyzed. There was too little data to estimate profiles for 8 occupations. Four occupations, AF, MA, NC, and LN, produced distinctive profiles. The difference between these occupations and the other clusters could not be summarized adequately by the cumulative hospitalization rate. However, these occupations represented only 0.9% of the total person years of observation.

Other evidence supports the conclusion that most USN occupations share a common general disease profile. Gorham, Garland, Helmkamp, and Gunderson (1987) examined engineering occupations during 2 periods, 1974 to 1979 and 1980 to 1983. The analyses were limited to White males, and diseases were classified using ICD-8. Disease rates were reported for 10 occupations: Machinist's Mate, Engineman, Machinery Repairman, Boiler Technician, Electrician's Mate, Interior Communications Electrician, Hull Technician, Gas Turbine Specialist, Patternmaker/Molder, and Fireman. Rates also were reported for all non-engineering occupations as a single group.

The profiles for 8 engineering occupations could be compared constructively. The profiles for Patternmaker/Molder (PM) and Gas Turbine Specialist (GS) were not included. There was too little data (i.e., <2,500 person years of observation for 1974 to 1979) for these occupations to estimate disease rates with reasonable precision.

The 8 engineering occupations included in the analysis had very similar profiles. Average similarity was r_s = .966 for 1974

Table 15. USN Disease Profiles: Engineering versus All Navy

	Engineering:		All Navy:		Relative Rate	
	74-79	80-83	74-79	80-83	74-79	80-83
Injury	209.52	181.56	157.00	153.74	1.33	1.18
Mental	106.26	101.95	92.54	103.53	1.15	.98
Digestive	77.57	69.46	76.62	784.03	1.01	.89
Musculoskeletal	84.08	100.96	75.27	99.02	1.12	1.02
Respiratory	62.35	47.24	54.48	47.47	1.14	1.00
Infectious	50.86	34.58	45.81	37.95	1.11	.91
Skin	44.00	33.51	36.56	31.32	1.20	1.07
Ill-Defined	36.54	24.40	34.15	30.00	1.07	.81
Genitourinary	27.50	24.49	31.48	30.72	.87	.80
Circulatory	25.08	18.03	29.71	26.55	.84	.68
Nervous	27.91	24.40	24.70	25.53	1.13	.96
Neoplasm	15.74	11.20	16.41	15.20	.96	.74
Endocrine	4.91	4.83	6.29	6.05	.78	.80
Blood	2.01	1.48	2.13	1.89	.94	.79
Total	774.32	678.04	683.15	687.38	1.13	.99

Note. Rates are given per 1,000 person-years at risk.

through 1979 and r_s = .966 for 1980 through 1983.[15] This degree of similarity was consistent with the earlier evidence that the pattern is the same for most USN occupations. The rates for these 8 occupations were combined to yield an overall engineering pattern. That pattern was virtually identical to the overall USN profile (1974-1979, r_s = .992; 1980-1983, r_s = .991). Table 13 gives the actual profiles.

The relative rate columns of Table 15 indicate some differences between engineering occupations and the overall U.S. Navy. Using a 10% difference in rates as a criterion, the higher Injury rate for engineering occupations was noteworthy. This elevation was offset by lower rates for engineering occupations in the Genitourinary, Circulatory, and Endocrine categories. The net result was a higher overall rate for engineering occupations during the 1974 to 1979 time period, but an equivalent rate for 1980 through 1983. The overall rate dropped from the first period to the last for engineering occupations (RR = 1.14), but not for other occupations (RR = 0.99).

The higher overall rate of hospitalizations for engineering occupations between 1974 and 1979 was the most striking element in the preceding comparisons. Engineering occupations had 96.27 more admissions per 1,000 years. Three disease categories accounted for 61.6% of the increase (Injury, 27.95 per 1,000

[15] If the two omitted occupations had been included in the analysis, the average similarity would have been rs = .899 for 1974 through 1979 and rs = .927 for 1980 through 1983.

person years; Respiratory, 15.11 admissions; Infectious, 16.28 per 1,000 person years). It has been noted previously that high rates for these 3 disease categories are typically encountered during periods of combat. The engineering profile, therefore, showed changes consistent with a shift from wartime to peacetime.

Musculoskeletal rates were noteworthy because they increased dramatically over time in both occupational groups. The absolute increase was 16.87 admissions per 1,000 person-years for engineering occupations and 23.75 admissions per 1,000 person years for other occupations. The relative risk of admission for musculoskeletal disease was RR = 1.20 for the engineering occupations and RR = 1.32 for the other occupations.

Hoiberg and Blood (1983) studied the USN officer population. The overall population was divided into aviators (n = 22,417), nonpilot aircrew members (n = 9,483), unrestricted line officers (n = 46,565), and staff corps (n = 55,593). Hospitalization rates for July, 1967, through 1980 were reported for 15 ICD-8 categories with coding based on ICD-8 categories.[16] Separate hospitalization rates were computed for aviators, aircrew personnel, staff personnel, and line personnel.

Table 16. USN Hospitalization Rates: Aviation versus Other Officers

	Aviators	Aircrew	Staff	Line
Digestive	80.5	81.1	44.9	44.0
Injury	67.9	77.5	32.4	29.0
Musculoskeletal	41.1	45.2	30.2	20.7
Genitourinary	30.9	32.0	22.4	14.3
Respiratory	28.7	42.0	29.0	17.5
Circulatory	25.1	14.8	20.0	15.2
Infectious	23.2	26.4	23.5	11.4
Ill-Defined	18.0	21.2	15.4	10.6
Neoplasm	14.7	14.6	8.9	9.6
Mental	14.0	23.6	20.2	12.0
Skin	13.7	20.0	9.4	9.1
Nervous	12.1	12.4	11.4	7.1
Congenital	4.4	4.0	3.2	2.0
Endocrine	3.9	4.8	2.9	2.9
Blood	1.2	1.6	1.0	.7
Total	379.4	421.2	274.8	206.1

Note. Rates per 10,000 person-years at risk as reported in Table I of Hoiberg and Blood (1983).

[16] Pregnancy and perinatal condition categories were excluded because the analysis was limited to males. The USN had too few women pilots during this period to accurately estimate rates for that population.

Patterns were highly similar across groups (median r_s = .943, range = .932 to .964) despite substantial differences in overall likelihood of hospitalization (Table 16). The rate of admissions for aviation personnel was ~67% higher than that for other Naval officers.

Comparisons to reference profiles produced two basic observations:

- Aviator profiles were more similar to peacetime profiles (median r_s = .789, range = .500 to .895) than to wartime profiles (median r_s = .342, range = .150 to .483).
- Aviator profiles appeared to be aviation specific. Among the peacetime profiles, the USN aviator profile most closely approximated the USAF profile (r_s = .778 to r_s = .895). The USN aviator profile was least similar to the general USN profile (r_s = .500 to r_s = .700).

Table 17. Comparison of Aviator Profiles to Reference Profiles

	Aviators	Aircrew	Staff	Line
Wartime				
World War I	.283	.333	.400	.217
World War II	.217	.267	.350	.150
Korean War	.350	.417	.450	.317
Vietnam War	.333	.400	.483	.350
Peacetime				
USAF	.778	.870	.895	.895
USA	.650	.800	.817	.817
USN	.500	.600	.583	.700

Note. Profiles consisted of 9 categories. Officer profiles based on Tables I through IV in Hoiberg and Blood (1983). Wartime reference profiles based on Table 1 of Hoeffler and Melton (1981). Peacetime reference profiles are the late profiles in Table 5 (p. 14).

Profile truncation affected the size of the similarity index. The values in Table 15 are limited to the 9 categories reported in Gardner et al. (1999). The similarities for the full 15 category wartime profiles were substantially higher (median r_s = .722, range = .625 to .825). These values still are less than those obtained with the truncated peacetime profiles. Any sizable increase in the similarities for peacetime profiles would make the aviator profiles approximate the USAF profile even more closely.

Table 18. Rate Ratios for Officer Categories

	Rate Ratio					
Of:	Aviator	Aviator	Aviator	Aircrew	Aircrew	Staff
To:	Aircrew	Staff	Line	Staff	Line	Line
Infective	.88	.99	2.04	1.12	2.32	2.06
Neoplasms	1.01	1.65	1.53	1.64	1.52	.93
Endocrine	.81	1.34	1.34	1.66	1.66	1.00
Blood	.75	1.20	1.71	1.60	2.29	1.43
Mental	.59	.69	1.17	1.17	1.97	1.68
Nervous	.98	1.06	1.70	1.09	1.75	1.61
Circulatory	1.70	1.26	1.65	.74	.97	1.32
Respiratory	.68	.99	1.64	1.45	2.40	1.66
Digestive	.99	1.79	1.83	1.81	1.84	1.02
Genitourinary	.97	1.38	2.16	1.43	2.24	1.57
Skin	.69	1.46	1.51	2.13	2.20	1.03
Musculoskeletal	.91	1.36	1.99	1.50	2.18	1.46
Congenital	1.10	1.38	2.20	1.25	2.00	1.60
Ill-Defined	.85	1.17	1.70	1.38	2.00	1.45
Injury	.88	2.10	2.34	2.39	2.67	1.12

Note. Ratios based on rates reported in Blood and Hoiberg (1983) and given in Table 14 above.

Cursory examination of the hospitalization rates in Table 14 will show that aviation personnel were distinguished from line and staff by higher hospitalization rates. Table 16 provides rate ratios showing that:

- Aviator/aircrew ratios varied from 0.59 (Mental) to 1.70 (Circulatory). However, the extreme ratios tended to be outliers. Other ratios ranged from 0.68 to 1.10.
- Staff/line ratios varied from 0.93 (Neoplasms) to 2.06 (Infectious). In this case, the highest ratio was an outlier. Excluding the two extreme ratios, the range was 1.00 to 1.68.
- Hospitalization rates in the aviation community were higher than those in the remainder of the officer population. Only 5 of 60 ratios were less than 1.00 compared to contrasts 16 of 60 that exceeded 2.00. All 4 ratios were greater than 2.00 for the Injury category. All 4 ratios were greater than 1.50 for Neoplasms and Digestive.

Aviation is a generally risky occupation. The rates for Injury, Digestive, and Neoplasms categories were particularly pronounced relative to officers in non-aviation specialties. The Digestive ratio also tended to be higher when the USN was compared to the USAF (Table 5). However, the USAF as a whole had fewer Injury hospitalizations than the USN as a whole. The similarity of the aviator profiles may seem straightforward, but the similarity of the actual work done may be less than it appears superficially. For example, USN personnel conduct a number of missions from aircraft carriers. This operational setting may provide a different mix of risks than flying from an

airfield. Still, the apparent convergence of aviation profiles, based partly on elevated frequency of hospitalizations for Digestive diseases, is a noteworthy point.

Hoiberg and Blood (1985, 1986) compared divers to matched controls. Sailors were divers if they had made at least 1 dive. Diving experience varied widely, but the average number of diving hours was 20.6 (range = 1 - 1046 hours) for officers. Diving experience was not reported for enlisted personnel. Matching was based on occupation, age, and length of service. Disease rates were based on hospitalization records with diagnoses coded according to ICD-8A categories. Hospitalization records covered the period 1968 through 1979.

Table 19. Hospitalization Profiles for USN Divers and Controls

	Officers[a]:		Enlisted[b]:	
	Divers	Control	Divers	Control
Musculoskeletal	78.2	51.8	124.6	116.1
Nervous	18.7	21.3	47.8	98.7
Stress-related[c]	15.5	35.0	34.5	34.7
Circulatory	20.4	6.1	25.4	34.9
Respiratory	16.3	22.1	22.2	35.6
Environmental[d]	2.4	.0	9.3	1.2
Total	151.5	136.3	263.8	321.2
Person-Years	5,841	6,432	73,272	55,824

[a]Rates from Hoiberg and Blood (1986) Table 1 (n = 1,973 in each group).
[b]Rates from Hoiberg and Blood (1985) Table 2 (n = 11,517 in each group).
[c]Mental disorders plus ulcers and diabetes mellitus
[d]Decompression sickness, effects of gas, drowning.

The 4 profiles were moderately similar (Table 20, see following page). The diver profiles were too dissimilar (r_s = .714) to clearly indicate an effect of diving status on the overall profile. This value was higher than the similarity of the 2 officer profiles (r_s = .429), but lower than the similarity of the 2 enlisted profiles (r_s = .771). The limited number of categories and the small total number of person-years at risk could have depressed these values.

Profile similarity was attributable primarily to the extreme categories. Musculoskeletal disorders were the most common reason for hospitalization in each group. Environmental exposures were the least common. In terms of relative risk, diving officers had excess rates of environmental problems (RR = ∞, rate = 0.00 in controls) and nervous and sensory system disorders than the matched controls (RR = 4.16), and

Table 20. Similarity of Diver and Control Profiles

	Officers[a]:		Enlisted[b]:	
	Divers	Control	Divers	Control
Officers				
Divers	1.000			
Controls	.429	1.000		
Enlisted				
Divers	.714	.714	1.000	
Controls	.829	.600	.771	1.000

[a]Profiles defined by rates from Hoiberg and Blood (1986) Table 1 (n = 1,973 in both officer group).
[b]Profiles defines by rates from Hoiberg and Blood (1985) Table 2 (n = 11,517 in both enlisted groups).

musculoskeletal disease (RR = 1.51). These differences were offset to some extent by lower rates of respiratory disease (0.44), so the overall risk of hospitalization for diving officers was only slightly higher than that of controls (RR = 1.11). Among enlisted personnel, divers had higher rates of environmental disorders (RR 7.75), but actually were less likely to be hospitalized overall (RR = .82). The lower cumulative risk could be attributed primarily to relatively low rates of stress-related disorders in the diver population (RR = 0.48). The cumulative rates produced RR = 0.97 when that category was removed.

Outpatient Care

Hospitalization is the tip of a medical care iceberg. Outpatient treatment is the body of that iceberg. Outpatient illnesses are mild relative to health problems that require hospitalization. However, outpatient treatment is so commonplace that the illness burden and time lost from work still are important.

The following examination of outpatient patterns parallels the topics covered for hospitalization data as far as possible. The parallels are limited by gaps in the data. Those gaps arise because studies of outpatient care traditionally have been more difficult to conduct than studies of inpatient care. Within the last few years, the military services have introduced computerized record keeping systems that will change this situation. The present review addresses what is known now.

General Population

Outpatient profiles for active duty personnel have been reported in the *Medical Surveillance Monthly Report*. For the period 1998 through 2003, outpatient rates have been reported for all active duty personnel combined across services. During this

period, the average stability coefficient was r_s = .979 (range = .950 to .997).

Table 21. Stability of Outpatient Disease Patterns

All Active Duty							
1998	1.000						
1999	.994	1.000					
2000	.997	.991	1.000				
2001	.982	.991	.985	1.000			
2002	.985	.974	.988	.968	1.000		
2003	.959	.974	.962	.988	.950	1.000	
Army Only							
1998	.988	.982	.991	.985	.985	.968	1.000
1999	.926	.921	.929	.906	.924	.988	.938	1.000

Note. Data for profiles were from annual rates reported by the U.S. Army Medical Center for Health Promotion and Preventive Medicine (2002, 2003, 2004). These reports are available on the internet at amsa.army.mil.

The stability coefficients for individual services may be somewhat lower than those for the overall active duty population. The profile stability for USA personnel was r_s = .938 from 1998 to 1999. When compared to the profiles for the all active duty personnel, the USA profile was highly similar in 1998 (r_s = .988), but somewhat dissimilar in 1999 (r_s = .926). USA personnel make up a large proportion of the total service population, so their health problems have a strong influence on the overall profile. The slight difference between the USA profile and total profile, therefore, suggests that illness patterns in other branches of the service differ from the USA profile. While the data to not make the case directly, these results would be consistent with the pattern of stable, but somewhat distinct profiles for each population.

Gender

An overview of gender differences in outpatient treatment rates can be developed by combining USA data from the *Medical Surveillance Monthly Reports* with USN data from Nice and Hilton (1990). The surveillance reports included ambulatory treatment rates for USA males and females for the years 1998 and 1999. Nice and Hilton (1990) reported outpatient treatment profiles for U.S. Navy personnel aboard 20 surface ships for October, 1988, through September, 1989. Ships included 7 destroyer tenders (AD), 2 repair ships (AR), 3 oilers (AS), 4 salvage ships (AO), and 4 submarine tenders (ARS). These ships were selected because their crews included females. The gender distribution for the 20 ships was 76% males (n = 10,869) and 24% females (n = 3,514).

The USA rates were based on reporting to a centralized computer database. The USN rates were estimated from sick call logs. Log entries were coded according to ICD-9 classifications for 16 disease categories.[17] Sick call rates were computed as the number of visits per 1,000 persons per month. A total of 62,671 patient visits translated to 440 visits per 1,000 per month for men and 780 visits per 1,000 per month for women. These rates converted to annual rates of 5,280 per 1,000 person years for men and 9,360 visits per 1,000 person years for women. The same conversion was applied to rates for individual categories to have numbers comparable to the *MSMR* rates.

Table 22. Outpatient Treatment Profiles for the USA and USN

	Males:			Females:		
	USA 98	USA 99	USN 89	USA 98	USA 99	USN 89
Infectious	187.9	170.8	608.4	374.9	341.7	83.6
Neoplasms	44.8	37.6	21.6	93.7	86.3	57.6
Endocrine	84.5	67.2	86.4	197.6	159.8	100.8
Blood	6.6	5.5	9.6	33.2	29.5	18.0
Mental	392.6	380.9	80.4	666.8	611.9	141.6
Nervous	307.3	278.7	271.2	546.3	510.5	452.4
Circulatory	83.4	71.3	43.2	104.7	98.8	51.6
Respiratory	397.7	255.6	865.2	868.0	778.4	870.0
Digestive	147.6	127.7	176.4	301.8	258.5	337.2
Genitourinary	80.0	69.4	56.4	810.4	682.5	782.4
Pregnancy	.	.	.	214.9	199.8	73.2
Skin	192.8	166.0	420.0	319.8	291.1	434.2
Musculoskeletal	1283.4	1192.0	422.4	2380.5	2238.1	580.8
Congenital	12.3	10.2	2.4	20.3	17.2	10.8
Ill-defined	310.2	295.0	202.8	914.5	872.5	559.8
Injury	752.1	663.7	1381.2	990.8	902.1	1381.2

Note. USA rates from Medical Surveillance Monthly Report. USN rates from Nice and Hilton (1990) cover the period October, 1988, through September, 1989. Navy rates, which were originally reported as visits per 1,000 persons per month, have been converted to visits per 1,000 persons per year.

Table 23. Similarity of Outpatient Treatment Patterns

Males						
USA 1998	1.000					
USA 1999	.964	1.000				
USN 1989	.846	.775	1.000			
Females						
USA 1998	.879	.886	.764	1.000		
USA 1999	.879	.886	.764	1.000	1.000	
USN 1989	.707	.696	.875	.874	.874	1.000

[17] Perinatal complications were not involved.

The pattern comparisons indicated both service and gender differences in outpatient patterns.

- The USA patterns were highly similar (males, r_s = .964; females, r_s = 1.000).
- The USN profile was only moderately similar to the USA profiles (males, average r_s = .811; females, average r_s = .874).
- Male and female profiles were moderately similar (average r_s = .815).

Outpatient rates differed substantially, but rate ratios for the two services would have questionable meaning. The rates for the USA are based on the entire population. The rates for the USN are based on a selected shipboard population. The fact that a decade intervenes between the rates also might affect the comparisons.

Table 24. Comparison of USN Shipboard Outpatient Profiles to Reference Profiles

	Females	Males
World War I	.804	.739
World War II	.845	.811
Korean War	.867	.793
Vietnam War	.829	.832
USAF	.146	.192
USA	.309	.583
USN	.224	.167

Note. Outpatient profiles based on rates reported by Nice and Hilton (1990). Based on 16 ICD categories for women (Perinatal Conditions omitted) and 15 categories for men (Pregnancy and Perinatal not relevant).

Similarity computations established that:

- o Male and female profiles were similar (r_s = .875).
- o Sick call profiles were much more similar to combat reference profiles (median r_s = .820) than to peacetime profiles (median r_s = .208; cf., Table 18).
- o The combat trio of injury, respiratory disease, and infection was clearly evident for both men and women (Table 19).

When the male-female patterns were examined in greater detail, the following points were important:

- The high rate of genitourinary problems for women was the primary for the less than perfect similarity of male and female profiles (5th vs. 12th, cf., Table 19).
- The female rate was higher in every disease category. The total sick call rate for women was 55% higher than for men.
- The rate ratios ranged from roughly equal (Respiratory, RR = 1.01) to a greater than 10-fold higher rate among women (Genitourinary, RR = 13.87). The rate for women was more than twice that for men for Congenital (RR = 4.50), Ill-defined (RR = 2.70), Neoplasms (RR = 2.65), and Nervous (RR = 2.08)

Table 25. Rate Ratios for Outpatient Treatment: Males and Females

	USA 98	USA 99	USN 88-89
Infectious	2.00	2.00	1.65
Neoplasms	2.09	2.30	2.67
Endocrine	2.34	2.38	1.17
Blood	5.03	5.36	1.88
Mental	1.70	1.61	1.76
Nervous	1.78	1.83	2.08
Circulatory	1.26	1.39	1.19
Respiratory	2.18	3.05	1.01
Digestive	2.04	2.02	1.91
Genitourinary	10.13	9.83	13.87
Skin	1.66	1.75	1.03
Musculoskeletal	1.85	1.88	1.37
Congenital	1.65	1.69	4.50
Ill-defined	2.95	2.96	2.76
Injury	1.32	1.36	1.17

Note. The USN rates from Appendix B of Nice and Hilton (1990), cover October, 1988, through September, 1989.

The rate ratios for gender showed higher rates (i.e., RR > 1.00) in every comparison. The difference was most pronounced for Genitourinary (RR > 9.82). The difference was also consistently >2.00 for the Ill-Defined (RR > 2.75) and Neoplasms (RR > 2.08) categories.

One element of the profile was surprising. The endocrine category ranked in the top 10 for males. This category consistently ranked lower in hospitalization profiles. The bulk of the visits that comprised this category dealt with obesity (males, 5.04 per 10,000 person weeks; females, 4.26 per 10,000 person weeks). These visits therefore were probably linked to weight control programs. Note that the female rate actually was lower than the male rate.

Geographic Region

Blood and Griffith (1990) reported outpatient data that approximate the regional hospitalization studies of Hermansen et al. (1990) and Pugh et al. (1989). These authors reported rates treatment rates for ships deployed to 3 regions, East Asia, Indian Ocean, and Europe.

Two points must be noted at the outset. First, temporal factors might be confounded with regional differences. The East Asia deployments took place between 1967 and 1973. The Indian Ocean and European deployments took place in 1985. Second, Blood and Griffith (1990) reported 9 rates (3 regions x 3 ship sizes). The information provided in their report has been reduced to 6 rates, 3 for ship sizes aggregating across regions (and time) and 3 for region aggregating across ship size. The sample sizes for specific combinations (e.g., small ships in the Indian Ocean) could be quite small. As illustrated previously, small samples would yield imprecise estimations of individual rates. The imprecision would be expected to reduce the size of the similarity coefficients. Smaller coefficients could give the appearance that profiles were specific to each region-time-ship size combination. This level of complexity is possible, but seemed unlikely given the hospitalization evidence. Pooling the measures across ship size and across region gave more stable profiles that should be adequate to detect their separate effects if any were present.

Disease rates were based on the ICD system. Rates were reported for 15 disease categories. The Pregnancy and Perinatal categories were not included.

Table 26. USN Outpatient Pattern Comparisons: Geographic Region and Ship Size

East Asia	1.000					
Indian Ocean	.960	1.000				
Europe	.910	.915	1.000			
Small	.992	.977	.915	1.000		
Medium	.968	.942	.932	.971	1.000	
Large	.979	.985	.918	.985	.964	1.000

Ship size did not affect the outpatient pattern. Similarity coefficients were very high (.964 ≤ rs ≤ .985) when patterns were compared across ship sizes. Geographical region produced highly similar patterns, although the European pattern did tend to differ somewhat from the East Asia (r_s = .910) and Indian Ocean (r_s = .915) patterns. The latter patterns were highly similar (r_s = .960). The lower ranking for Genitourinary (10^{th} vs. 5^{th}) was the primary difference between the European patterns for East Asia and the Indian Ocean.

Table 27. Disease Rates by Region: USN Personnel Deployed Aboard Ships

	East Asia[a]	Indian Ocean[b]	Europe[b]
Infectious	645.74	887.07	488.29
Neoplasms	3.00	2.12	.00
Endocrine	7.20	2.83	62.40
Blood	3.60	4.24	.00
Mental	71.42	62.86	87.59
Nervous	150.33	150.43	104.01
Circulatory	10.20	41.67	55.84
Respiratory	995.62	817.15	1209.77
Digestive	182.14	89.70	170.79
Genitourinary	452.50	391.98	83.21
Skin	481.90	621.51	528.80
Musculoskeletal	181.84	416.70	342.68
Congenital	1.50	.00	.00
Injury	544.92	495.80	789.36

Note. Rates given as number of visits per 1,000 per year based on Tables I, II, and III of Blood and Griffith (1990). Those tables report rates for small, medium, and large ships in each region. The pooled rates were computed for this review from data in the tables.

[a]From 1967 through 1973
[b]During 1985

The relative rates for different disease categories generally were comparable across regions. High rates for Genitourinary disease outside of Europe were the primary exception to this generalization. The relative rates for this disease category were RR = 5.44 for East Asia and RR = 4.71 for Indian Ocean. Other than the Genitourinary rates, the most noteworthy difference was that Musculoskeletal rates in East Asia were approximately half the rates in the Indian Ocean (RR = 2.28) and Europe (RR = 1.94).

Several European rates were high relative to East Asia and Indian Ocean rates. However, the low absolute magnitude of the rates involved limits the importance of those differences for health care planning. For example, the European rate for Endocrine was 8.7 times that for East Asia and 22.0 times that for the Indian Ocean. However, the absolute rate was only 62.40 per 1,000 person years in Europe. By comparison, the rate in Europe for Respiratory was 1209.77 per 1,000 person years while that for Injury was 789.36 per 1,000 person years. The high relative rates for Endocrine, therefore, reflected primarily the virtual absence of diagnoses in this category for personnel deployed outside of Europe rather than a high rate (in absolute terms) in USN personnel stationed in Europe.

Table 28. Comparison of USN Outpatient Profiles for Deployed Shipboard Populations to Reference Hospitalization Profiles

	WW I	WW II	Korea	Vietnam
East Asia	.857	.900	.968	.843
Indian Ocean	.772	.845	.887	.801
Europe	.742	.828	.817	.892
Small	.810	.860	.944	.808
Medium	.821	.861	.936	.832
Large	.814	.871	.929	.818
Combined	.836	.882	.957	.832

Note. Based on 15 disease categories. Pregnancy and Perinatal were omitted.

Blood and Griffith's (1990) outpatient profiles were more similar to wartime reference profiles than to peacetime reference profiles Table 23). The similarity coefficients for comparisons to wartime hospitalization profiles ranged from r_s = .742 to r_s = .968 (Table 22). Similarity coefficients for peacetime hospitalization profiles ranged from r_s = -.200 to r_s = .683. These distributions did not even overlap.

Table 29. Comparison of USN Profiles to Peacetime Profiles

	USAF	USA	USN
East Asia	.192	.517	-.100
Indian Ocean	-.017	.377	-.159
Europe	.285	.617	.233
Small	.159	.533	-.083
Medium	.285	.683	.100
Large	-.092	.383	-.200

Note. Based on 9 disease categories.

The reasons why the outpatient patterns were similar to wartime profiles and differed from peacetime patterns were clear. The most common diseases in the wartime hospitalization profiles ranked highly in the outpatient patterns. The Respiratory, Infectious, and Injury categories were the among the top 4 disease categories in each regional pattern. Skin disease was 3rd or 4th in each case. Relatively low rankings for the Digestive (6th) and Mental (9th) categories also differentiated these outpatient patterns profiles from the peacetime hospitalization reference patterns.

Table 30. Disease Rates by Ship Size

	Small	Medium	Large
Infectious	935.75	617.91	586.45
Neoplasms	.00	2.31	3.21
Endocrine	4.03	6.94	20.3
Blood	2.02	.00	4.01
Mental	91.36	55.54	66.02
Nervous	149.13	171.26	137.12
Circulatory	10.08	71.74	26.20
Respiratory	1606.16	1011.34	735.60
Digestive	302.29	222.17	91.95
Genitourinary	421.19	238.37	376.62
Skin	597.19	453.60	503.59
Musculoskeletal	321.10	314.74	239.23
Congenital	.00	4.63	.80
Ill-Defined	139.05	166.63	103.18
Injury	557.56	786.86	553.04

Table 31. Relative Rate by Ship Size

	Ratio	
Of:	Small	Medium
To:	Medium	Large	Large
Infectious	1.51	1.60	1.05
Neoplasms	.00	.00	.72
Endocrine	.58	.20	.34
Blood	.[a]	.50	.00
Mental	1.64	1.38	.84
Nervous	.87	1.09	1.25
Circulatory	.14	.38	2.74
Respiratory	1.59	2.18	1.37
Digestive	1.36	3.29	2.42
Genitourinary	1.77	1.12	.63
Skin	1.32	1.19	.90
Musculoskeletal	1.02	1.34	1.32
Congenital	.00	.00	5.77
Ill-Defined	.83	1.35	1.61
Injury	.71	1.01	1.42

[a]Ratio could not be computed because the estimated rate was 0.00 for medium sized ships.

Rates varied widely (RR = 0.00 to RR = 5.77), but much of this variation could be attributed to the generally low rates of within some categories. Limiting attention to the categories with rates > 1.00 for at least 2 ship sizes, the following trends would be important:

- Infectious and Respiratory rates declined with ship size.
- Small ships had the highest rate in the Skin disease category; rates for medium and large ships were approximately equal.

- o Medium size was associated with higher Injury rates.
- o Medium size was associated with lower Genitourinary rates.

Given the lack of a priori hypotheses, any of these findings might be attributed to chance. However, if size operated as a causal factor in some way, one would expect consistent patterns of increase or decrease. This perspective would direct the most attention to the increasing rates for Infectious and Respiratory diseases. Many respiratory diagnoses involve infections (e.g., pneumonia), so both trends could indicate greater probability of spreading infections aboard smaller ships. This would be a reasonable finding if being a member of a smaller crew increases the likelihood of interacting with one or more infected individuals when a pathogen is present.

Combat Status

Defining the effects of combat on health care requirements of military populations clearly is an important issue. This topic has received attention in several studies of outpatient health care. Rahe, Mahan, Gunderson, & Arthur (1970) reported rates for 3 cruisers during deployments. The cruisers were chosen to contrast combat deployment to Vietnam (Cruisers 1 and 2) with a peacetime Mediterranean deployment (Cruiser 3). However, the outbreak of hostilities between Arabs and Israelis during the deployment blurred this distinction. Outpatient diagnoses were classified into 13 categories based on a schema developed within the research project. The classifications were closely linked to functional system (e.g., respiratory, dermatological, aural, cranial). Ship patterns generally were similar (Cruisers 1 and 2, r_s = .847; Cruisers 1 and 3, r_s = .851; Cruisers 2 and 3, r_s = .935). The major difference between the profiles was a lower proportion of gastrointestinal problems aboard Cruiser 1. That category ranked 7th aboard Cruiser 1 and 3rd or 4th aboard the other cruisers.

Blood and Nirona (1990) provided outpatient profiles that can be used to examine profiles related to ship type and combat status. The profiles covered 11 ICD categories for USN personnel during the Vietnam War. Profiles were available for 2 ship classes: Aircraft Carriers (AC) and Destroyer/Frigates (DF). AC data were available for 6 years during the conflict (1967-1972) and for the cease-fire year (1973). DF data were available for 2 years during the conflict (1971, 1972), the cease-fire year (1973), and one post-conflict year (1975).

Combat status affected the patterns (Table 32). Combat patterns were only moderately similar to the 1973 patterns (AC, r_s = .627; DF, r_s = .840). Stability was much higher for the post combat DF profiles (r_s = .964).

Ship type effects on patterns were more difficult to assess. The patterns were dissimilar during the combat period (r_s = .626), but quite similar in 1973 (r_s = .936). These results suggest that the effects of combat may be closely linked to the actual activities and operational tempo of different units.

Three disease categories accounted for the effect of combat on the AC pattern. Injury ranked 1st during combat and 5th following the conflict. Genitourinary disorders ranked 7th during combat and 1st following combat. Musculoskeletal disorders ranked 10th during combat and 6th afterward. The only substantial change in the DF pattern was that Skin disease moved from 7th during combat to 3rd after combat.

Table 32. Profile Similarity for Ship Types in Vietnam War

Aircraft Carrier					
1967-1972	1.000				
1973	.627	1.000			
Destroyer/Frigate					
1971-1972	.626	.813	1.000		
1973	.727	.936	.840	1.000	
1975	.736	.873	.863	.964	1.000

Note. Based on data reported in Blood and Nirona (1990).

Table 33. Comparison of Vietnam Era Ship Type Disease Patterns to Reference Patterns

	Aircraft Carrier:		Destroyer/Frigate:		
	1967-1972	1973	1971-1972	1973	1975
World War I	.700	.564	.804	.636	.609
World War II	.800	.527	.758	.627	.609
Korean War	.818	.855	.904	.900	.882
Vietnam War	.655	.291	.511	.473	.436
USAF	.042	-.042	.159	.142	.226
USA	.367	.333	.467	.500	.600
USN	-.117	.217	.133	.117	.017

The outpatient patterns clearly were more similar to wartime hospitalization patterns:

- During the combat period, the similarity coefficients for comparisons to wartime reference profiles ranged from r_s = .511 to r_s = .904 (median r_s = .779). The coefficients for comparisons to peacetime hospitalization patterns ranged from r_s = -.117 to r_s = .467 (median r_s = .088). The clear convergence with the combat reference pattern was evident

in the fact that there was no overlap in the distributions of similarity coefficients.

- Outpatient treatment was more similar to combat hospitalization profiles even after combat ended. Directing attention to 1973, the similarity coefficients for comparisons to wartime reference profiles ranged from r_s = .291 to r_s = .900 (median r_s = .578). The coefficients for comparisons to peacetime hospitalization patterns ranged from r_s = -.042 to r_s = .500 (median r_s = .180).

Direct examination of the outpatient rates provides insight into the disease pattern. Some noteworthy points were:

- Respiratory rate, which ranked 1st or 2nd in every profile, contributed heavily to the similarity to combat profiles.
- Total treatment rates were lower during combat. For 1973, the difference was 9.5% for AC and 69.3% for DF.
- Post-combat increases in Infectious, Genitourinary, Skin, and Musculoskeletal admissions accounted for 1,048.85 more illnesses per 1,000 person years aboard AC and 618.46 more illnesses per 1,000 person years aboard DF.
- Respiratory rates provided a striking contrast between AC and DF. Following combat, the AC rate for this category dropped by 21%, while the DF rate increased by 117%.

Table 34. USN Outpatient: Vietnam Shipboard Rates

	Aircraft Carriers		Destroyers/Frigates		
	1967-72	1973	1971-72	1973	1975
Respiratory	748.52	591.68	1278.56	2774.53	2819.68
Infectious	284.08	536.61	1285.50	1525.25	2450.06
Injury	753.03	266.22	548.95	646.18	823.73
Genitourinary	131.89	660.95	458.61	689.26	614.63
Skin	347.20	504.06	180.67	864.27	1083.52
Musculoskeletal	34.95	245.35	354.38	405.21	517.47
Nervous	181.49	111.83	180.67	284.05	192.20
Digestive	151.06	94.30	201.51	374.25	418.20
Ill-Defined	111.60	25.04	41.69	49.81	57.03
Mental	63.13	47.57	34.74	161.54	88.71
Endocrine	14.65	5.01	41.69	26.92	154.18
Total	2821.59	3088.91	4606.97	7801.27	9219.41

Note. Based on data from Table I and Table II of Blood and Nirona (1990). Rates reported as cases per 1,000 strength per day were converted to rates per 1,000 strength per year for comparison to other profiles.

Other things equal, combat injuries increase the overall disease burden. Finding that overall disease rates were equal (AC) or lower (DF) during combat therefore was surprising. Blood and Nirona's (1990) suggest that decisions about seeking treatment are an important factor in this result. During combat, sailors might seek care only for relatively severe illness. The

fact that the proportion of visits for injuries dropped after combat ceased is consistent with this view. If disease rates are affected by decisions that change during combat, peacetime rates will be a poor guide to planning for combat medical care.

The rate ratios in Table 35 direct attention to differences as a function of combat status and ship type. The most important trends were:

- Combat status was associated with consistent rate differences in several categories
  - Rates for Infectious, Genitourinary, Skin, Musculoskeletal decreased during combat.
  - The Endocrine rates were the only ones that increased during combat aboard both ship types.
  - Effects on Injury and Mental, both of which might have been expected to increase, were inconsistent across ship types.
- The only ship type effects that were reliable over time were lower disease rates aboard AC for Infectious, Endocrine, Respiratory, Digestive, Genitourinary, and Musculoskeletal.

Table 35. Disease Rate Ratios: Combat Status and Ship Type

	Combat[a]		Ship Type[b]	
	AC	DF	During	After
Infectious	.53	.84	.22	.35
Endocrine	2.93	1.55	.35	.19
Mental	1.33	.22	1.82	.29
Nervous	1.62	.64	1.00	.39
Respiratory	1.27	.46	.59	.21
Digestive	1.60	.54	.75	.25
Genitourinary	.20	.67	.29	.96
Skin	.69	.21	1.92	.58
Musculoskeletal	.14	.87	.10	.61
Ill-Defined	4.46	.84	2.68	.50
Injury	2.83	.85	1.37	.41

Note. Based on rates reported by Blood and Nirona (1990). AC = Aircraft Carrier; DF = Destroyer/Frigate.

[a]Combat = Ratio of combat rate to 1973 rate for the indicated ship type.
[b]Ship Type = Ratio of AC rate to DF rate for the specified period.

Ship Type

The preceding discussion of combat effects raises an issues that is unique to the USN and USMC personnel assigned to Navy ships. Does ship type affect disease patterns? That topic has been investigated in several studies.

Bohnker, Sherman, and McGinnis (2003) reported outpatient rates for 15 diseases categories defined by the Joint Chiefs of Staff syndromic criteria. Reports covered 217,972 person-weeks (4,192 person years) of observation on 233 ships of the USN Fifth Fleet. Rates were given for supply ships (SS), larger surface ships (i.e., cruisers, destroyers, and frigates, CDF), aircraft carriers (AC), and amphibious ships. The CDF, AC, and Amph profiles were highly similar (r_s > .929). The SS profile was somewhat different even though it was highly similar to the CDF profile (r_s = .936).

The pattern similarity coefficients for different ship types indicated substantial convergence across ship types. This assertion is appropriate even though they were only moderately high compared to some other similarities noted in this review. The similarities must be interpreted in light of the fact that each profile was based on relatively limited data. The period of observation amounted to 2,691 person-years for AC, 1,207 for CDF, 161 for Amphibious, and 132 for SS. This modest volume of data suggests that rate estimates may be more susceptible to chance variation than some of the other rates covered in this report. With this qualification in mind, the similarities that range from r_s = .861 to r_s = .946 with a median of r_s = .933 are high relative to other profile comparisons reported here.

Table 36. Profile Comparisons Across Ship Type

	SS[a]	CDF[b]	AC[c]	Amphibious
SSS[a]	1.000			
CDF[b]	.936	1.000		
AC[c]	.861	.946	1.000	
Amph[d]	.877	.930	.932	1.000

Note. Based on rates reported in Bohnker et al.'s (2003) Table II.
[a]Supply Ships
[b]Cruiser/Destoyer/Frigate
[c]Aircraft Carrier

Tansey, Wilson, and Schaefer (1979) examined the health of USN submariners. This population might have an atypical profile because of extensive selection screening to ensure good physical and mental health. This screening is essential because submarines perform extended patrols. Any significant health problems arising during these deployments could severely compromise mission success.

Tansey et al. (1979) compared illness rates of submarines to those of surface ships. The medical sections of 885 submarine

patrol reports provided the submarine data. Patrols were divided into two periods, 1963 to 1967 and 1968 to 1973, represented by 3,240,000 person-days (8,871 person-years) of observation and 4,410,000 person-days (12,074 person-years) of observation, respectively. Surface ship rates were based on data from a continuous 7-8 month deployment of destroyers, destroyer escorts, and guided missile destroyers. The data for the surface ships represented 1,215,918 person-days (3,329 person years) of observation from "... a typical deployment of the vessels described" (Tansey et al., 1979, p. S219). Disease classifications were based on a standard schema for submarine patrol reports (see Table 28 for categories). Analysis indicated:

- The submarine pattern was stable over time (r_s = .919).
- The pattern was stable even though the submarine disease rate dropped 31% (Difference = .083; RR = .69) from the first period to the second.
- Most of the drop in the overall rate was associated with decreases in resulted respiratory disease (Difference = .043, RR = 0.39), trauma (Difference = .014, RR = 0.72), and gastrointestinal disease (Difference = .015; RR = 0.70) rates.
- Both submarine patterns differed from the surface ship pattern (1963-1967, r_s = .758; 1968-1973, r_s = .808).
- The major pattern differences were higher rankings for cranial injuries (6^{th} or 7^{th} vs. 10^{th}) and lower rankings for miscellaneous problems (10^{th} vs. 7^{th}) in the submarine population.
- Both submarine rates were substantially lower than the 1973 surface rates (1963-1967, RR = 0.62; 1968-1973, RR = 0.42).

Table 37. Disease Rates for Submarine and Surface Ships

	Submarines:		Surface
Category	63-67	68-73	73
Respiratory	.070	.027	.130
Trauma	.050	.036	.100
Gastrointestinal	.050	.035	.090
Dermal	.020	.022	.020
Infection	.020	.014	.060
Genitourinary	.020	.014	.010
Systemic	.010	.010	.004
Cranial	.020	.013	.003
Neuropsychiatric	.004	.012	.006
Miscellaneous	.003	.001	.010
Total	.267	.184	.433

Population Differences

Shaw, Hermansen, Pugh, and White (1991) reported disease patterns for USN and USMC personnel during Operation Desert

Shield/Desert Storm. The data used to construct the patterns were obtained from 2 field hospitals for the period from September, 1990 to March, 1991. Diagnoses were classified according to ICD-9. Thirteen of 17 disease categories were represented in the data.

This information has been placed in this part of this review because most medical care involved outpatient treatment. Four of every 5 cases (81.1%) were returned to full duty. About 1 in every 8 patients was returned to his or her unit with a duty restriction (light duty, 8.0%; no duty, 4.9%). The remaining 6% were hospitalized (1.5% evacuated; 4.5% admitted to the field hospital).

Table 38 shows the disease profiles expressed as the percentage of cases in each category. The USN and USMC profiles were moderately similar (r_s = .870). The largest difference was that Genitourinary ranked 4th for USMC personnel and 9th for USN personnel.

Table 38. Distribution of Cases by Disease Category

	Branch of Service:	
Category	USN	USMC
Respiratory	28.83	9.22
Injury	21.17	37.65
Ill-Defined	13.22	7.84
Skin	9.05	10.39
Infectious	5.86	6.47
Nervous	5.77	5.88
Musculoskeletal	5.37	6.47
Digestive	3.98	5.49
Genitourinary	2.88	8.04
Circulatory	1.69	2.16
Mental	.99	.20
Neoplasms	.60	.20
Endocrine	.60	.00

Note. Table entries are percentages of total disease cases in each of 13 disease categories computed from data in Table 1 of Shaw, et al. (1991). Percentages differ from figures in that table because categories of "No Diagnosis" and "Supplementary Classification" were dropped from the computations. Four categories, Pregnancy, Perinatal, Congenital, and Blood did not appear in the table.

Table 39. Profile Similarities for Operation Desert Storm/Desert Shield

	USN[a]	USMC[a]	USN[b]	USMC[b]
Wartime				

World War I	.702	.556	.600	.368
World War II	.834	.683	.767	.469
Korean War	.680	.708	.467	.527
Vietnam War	.812	.675	.717	.536
Peacetime				
USAF			-.343	-.097
USA			-.017	.025
USN			-.050	-.050

[a]Based on 13 categories common to the reference pattern and the Shaw et al. (1991) patterns.
[b]Based on 9 categories common to the reference pattern and the Shaw et al. (1991) patterns.

Note. Operation Desert Storm/Desert Shield profiles were based on data from Shaw et al. (1991; see Table 38 above). Peacetime profiles were defined by hospitalization rates reported in Gardner et al. (1999, see Table 5, p. 14). Wartime profiles were defined in Table 1 of Hoeffler and Melton (1981).

Reference pattern comparisons raised two points:

- Desert Storm patterns tended to be unique. The patterns were weakly similar to past wartime patterns (median r_s = .532, range = .368 to .767). They were even less similar to peacetime profiles (median r_s = -.050, range = -.343 to .025).
- Despite the low overall similarities for the Desert Storm patterns, the Injury and Respiratory categories still were among the most common problems. These categories were the primary sources of similarity to prior wartime patterns.
- Relatively low rankings for Infectious (USN = 5^{th}, USMC = 7^{th} vs. Reference = 2^{nd} or 3^{rd}) and Digestive (USN = 8^{th}, USMC = 9^{th} vs. Reference = 3^{rd} - 5^{th}) and high rankings for Skin (USN = 4^{th}, USMC = 2^{nd} vs. Reference = 7^{th} - 10^{th}) were the most important sources of differences between the Desert Storm patterns and wartime hospitalization patterns.

The Desert Storm patterns underscore the risks associated with extrapolation. The distinctive character of these patterns was striking, but methodological considerations cannot be discounted as a potential explanation. The amount of data available to define these patterns was relatively modest compared to the volumes of data that generated some of the other findings. The absence of some disease categories also affected the findings. Similarities no doubt would have increased if the absence of Endocrine and Blood diagnoses had been treated as true zeros rather than omitting them.

Summary

Outpatient treatment patterns are broadly similar to combat hospitalization profiles. Distinctive patterns can be identified for men and women. Rates generally are higher for women. Factors such as ship type and combat status may have little effect on these profiles. The Desert Storm evidence was a reminder that patterns differ across specific operational settings.

General Summary

Looking Back

The evidence summarized here provides some general insight into the health of U.S. military personnel. In this regard, the major findings were:

- The wartime hospitalization profile is different than the peacetime profile. The Injury, Infectious, and Respiratory categories are among the most important during wartime. The key elements of the peacetime pattern are Digestive, Pregnancy, and Musculoskeletal. Combat patterns appear to be affected by a discretionary element in seeking health care. Studies that compared the same population during combat and following combat found consistent increases in some categories after the cessation of hostilities (e.g., Blood & Nirona, 1990).

- Trend analysis is of limited value for forecasting future disease rates. Some clear trends were evident in the 1980s and early 1990s, but attempts to generalize were limited by several factors. The trends did not apply to the USA during that period. Extrapolation from the trends produced unrealistically low rates when extrapolated to 2004. When the time period covered by the analysis was extended for the USA, the results were very different. Finally, trends were not evident in USN and USMC data from 1980 through 1984. More sensitive methods of estimating future patterns from past rates are going to be needed.

- Patterns can remain relatively constant even when trends are present. The overall hospitalization rate for USN and USMC personnel dropped ~88% from ~800 per 1,000 per year to ~100 per 1,000 per year since the beginning of the 20th century (Hoeffler & Melton, 1981). Despite this, the disease pattern has remained relatively constant.

- General profiles are of limited value. Analyses that compared population patterns over time typically had 2 distinctive characteristics. Patterns were stable over time within populations. Patterns differed substantially between

populations even when the comparison was limited to a single year. Taken together, these points suggest that patterns are population specific.

- Profiles differ between men and women. Pregnancy, which is the primary hospitalization category for women, is not relevant to men. Among the categories that are relevant to both sexes, Genitourinary category ranks much higher for women. This difference was evident for outpatient treatment as well as for hospitalization. Finding differences between men and women is not at all surprising, but noting those differences does serve as a reminder that gender differences are a factor in understanding disease rates in military populations.

- Some attributes that define distinctive populations can be identified. These include obvious attributes such as combat status and gender. However, expected differences were absent in some cases. At least within the USN population, occupational differences were missing. The difference between occupations appeared to be general elevation of the profile more than the specific pattern of diseases. Ship type, a variable unique to the USN population, may affect disease patterns, but the overall evidence was inconsistent.

- Outpatient disease patterns were somewhat more similar to combat hospitalization patterns than to peacetime hospitalization patterns. However, these patterns were only moderately similar to the combat pattern. Outpatient patterns therefore appear to define a third category of disease pattern. These illnesses obviously are mild relative to those that require hospital admissions. However, these mild health problems are much more frequent than severe health problems. The relative contribution of outpatient disease to personnel losses during any given year may be greater than it may appear at first. Systematic analysis of duty time losses would be needed to fully evaluate this issue.

What are the implications of these somewhat banal general statements? These observations are reminders of one important point. Military disease patterns are affected by a wide range of factors. These factors include, but are not limited to, recruitment, retention, geopolitical events, economics, trends in medical care (e.g., managed care), and so forth. Trends are relevant to disease forecasting when it is reasonable to expect the underlying conditions to remain constant. It is not reasonable to expect the complex mixture of factors that affect the health of military populations to remain constant for any extended period of time. Neither is it reasonable to expect all

of the factors to change at the same rates and in directions that have the same effect on military health. The best approach to dealing with this complex set of factors would be forecasting models akin to those found in econometrics. Factors that affect the rates would be identified and modeled with trend forecasts based on the models, including changes in indicators. If complex models of this sort exist and are in use at present, but, the published literature does not document the practice.

Logical analysis will not suffice. Logic coupled with past experience can generate useful hypotheses. However, those hypotheses must be tested. For example, logic might lead one to expect that combat will increase a variety of disease rates. Certainly this expectation should hold for injuries, and it usually does. However, injuries did not increase among aircraft carrier personnel in Vietnam (Blood & Nirona, 1990). In fact, the general picture for Vietnam data suggested that the prevalence of some types of disease may decrease during wartime because treatment can be delayed.

The available evidence does not provide tests of all of the meaningful hypotheses that might be of interest. In fact, the motivation behind most studies appears to have been simple description. Description is a useful starting place, but the results leave inferential gaps when attempting to construct a viable model of the health of military populations. Given this type of information, all of the general assertions offered above should be considered open to question. For example, assertions about combat effects are based on general populations within a combat theater. Hoeffler and Melton's (1981) profiles clearly apply to wartime, but personnel did not even have to be in a combat theater to be included in their profiles. Dlugosz et al. (1999) provides an example of more detailed analyses that could better define combat effects.

The assertion that occupational profiles differ primarily with respect to overall disease rate is another example of a debatable conclusion. Vickers (1998) found that differences in injury and musculoskeletal disease rates were linked to occupational differences in physical demands. Other disease rates were not related to those demands. This specificity of association is at odds with the conclusion that occupation affects the overall profile rather than specific elements. Note also that many occupations have similar levels of physical demands. Thus, occupation may not be a very useful construct for predicting disease patterns when occupation is considered in the abstract. Studies of more specific factors such as exposure to physical demands and environmental hazards are likely to be essential to understanding disease patterns.

The preceding comments on the effect of combat and occupation on disease rates are instances of a general limitation

of the available evidence. The evidence is not sufficiently detailed to understand the factors contributing to disease well enough to predict future patterns. One aspect of the lack of detail is particularly striking. The scope of the available information for the USAF and USA populations is more limited than that for the USN and, to a lesser extent, the USMC. As a consequence, the generality of the inferences made here cannot be evaluated as well as would be desirable.

The available evidence also suffers from methodological limitations. Epidemiological research generally includes adjustments for factors that might bias estimates of disease rates and produce mistaken inferences. For example, disease rates are commonly adjusted for age and gender. Changes in policy meant that more women have joined the services in recent years. It is reasonable to wonder how much this trend contributed to the increase in pregnancy rates during the 1980s and early 1990s. Assuming that the upward trend remained after introducing this adjustment, adjustments for age might be a reasonable next step. Perhaps better career opportunities are keeping more women in the service long enough to start families. In short, the rising pregnancy trend is of uncertain significance when taken out of context. In general, it is arguable that changes in the composition of the services following the switch to an all-volunteer force could affect both general health trends and specific disease patterns in more narrowly defined populations (e.g., sailors in the Indian Ocean).

The assumption that disease profiles are the product of complex causal processes is reasonable. Those processes quite probably range from societal trends (e.g., changes in smoking habits of the general population) to recruitment and training practices in the military to administrative regulations (e.g., combat assignments for women) to operational tempo to climatic conditions, exposure to environmental hazards, and geopolitical factors. Given this complexity, the diligent efforts of past researchers[18] provide insight into disease patterns and context for more detailed attacks on the questions that currently are unanswered.

Thinking Ahead

The limitations of current knowledge indicate the need for improved epidemiological tools for planning health care delivery to military populations. Today is a propitious time to direct

[18] The effort required to extract evidence from sick call logs or even from huge files of punched computer cards is tremendous. The statement that the current picture is incomplete is in no way a criticism of effort that was expended to provide even this picture. This review is an attempt to increase the utility of that hard won knowledge by putting it in a compact form to use as leverage for new work.

attention to this point. The work required to conduct a single study of a general population has been a major barrier to better medical intelligence in the past. Current applications of information technology are reducing the size of this barrier. For example:

- Large relational databases are being constructed that can be powerful tools for developing a better understanding of military health issues. Examples include the databases used to support the *Medical Surveillance Monthly Report* and the Defense Manpower Data Center-Combined Health and Military Personnel database (DMDC-CHAMPS) at the Naval Health Research Center. Both of these databases are examples of triservice cooperation as both include data from all branches of the military.

- Computerized systems that can record the medical treatment of individual patients in some detail are being developed. The systems are intended primarily to help hospital administrators implement managed care practices, but the information can provide insight into treatment patterns.

- Detailed databases are being constructed to assess specific problems. A tri-service program is developing a Combat Trauma Registry. A mishap reduction system is being constructed that integrates existing accident databases that provide detailed descriptions of accidents with health care and career history data.

- Systems designed to provide real-time monitoring of the health status of operational personnel are being developed and implemented. These systems include standard formats for collecting and storing medical care data (e.g., Joint Medical Work Station, JMeWS) and methods of detecting epidemics of any type of disease (e.g., Medical Data Surveillance System, MDSS). The primary purpose of these systems is to help medical planners monitor the medical status of operational units. The data gathered for that purpose can be used retrospectively to investigate operational health care in the field.

The full potential of these new systems will not be realized by business as usual. This approach yields studies of convenience samples, piecemeal surveillance of specific health problems, just-in-time analyses of specific situations, and after-the-fact analyses of unique events. Systematic exploitation these opportunities is a prerequisite for converting data into information. This review of past work may help provide a framework for planning that systematic exploitation. The available evidence provides a general sketch with many blank spaces. The barriers to filling in the blanks are being lowered

dramatically. The next step is to exploit the resulting opportunity.

References

Blood, C. G., & Aboumrad, T. L. (2001). A comparison of postdeployment hospitalization incidence between active duty Vietnam and Persian Gulf War veterans. *Military Medicine, 166*(7), 648-655.

Blood, C. G., & Gauker, E. D. (1993). The relationship between battle intensity and disease rates among marine corps infantry units. *Military Medicine, 158*(5), 340-347.

Blood, C. G., & Griffith, D. K. (1990). Ship size as a factor in illness incidence among U.S. Navy vessels. *Military Medicine, 155*(7), 310-314.

Blood, C. G., Jolly, R., Gauker, E. D., & Pugh, W. M. (1994). Comparisons of casualty presentation and admission rates during various combat operations. *Military Medicine, 159*(6), 457-461.

Blood, C. G., & Nirona, C. B. (1990). Outpatient illness incidence aboard U.S. Navy ships during and after the Vietnam conflict. *Military Medicine, 155*(10), 472-476.

Bohnker, B. K., Sherman, S. S., & McGinnis, J. A. (2003). Disease and nonbattle injury patterns: afloat data from the U.S. Fifth Fleet (2000-2001). *Military Medicine, 168*(2), 131-134.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second ed.). Hillsdale, NJ: Erlbaum.

Dlugosz, L. J., Hocter, W. J., Kaiser, K. S., Knoke, J. D., Heller, J. M., Hamid, N. A., Reed, R. J., Kendler, K. S., & Gray, G. C. (1999). Risk factors for mental disorder hospitalization after the Persian Gulf War: U.S. armed forces, June 1, 1991 - September 30, 1993. *Journal of Clinical Epidemiology, 52*(12), 1267-1278.

Gardner, J. W., Amoroso, P. J., Grayson, J. K., Helmkamp, J., & Jones, B. H. (1999). Hospitalization due to injury: inpatient medical records data. *Military Medicine, 168*(8, Supplement), 5.1-5.143.

Gorham, E. D., Garland, F. C., Helmkamp, J. C., & Gunderson, E. K. E. (1986). *Disease and injury in U.S. Navy engineering occupations* (Technical Report 87-8). San Diego, CA: Naval Health Research Center.

Gunderson, E. K. E., & Colcord, C. (1982). *Health risks in naval occupations: an overview* (Technical Report 82-1). San Diego, Ca: Naval Health Research Center.

Hermansen, L. A., White, M. R., Pugh, W. M., & Shaw, E. K. (1990). *Disease and non-battle injury rates for marine corps enlisted personnel during peacetime* (Technical Report 90-10). San Diego, CA: Naval Health Research Center.

Hoeffler, D. F., & Melton, L. J., III. (1981). Changes in the distribution of navy and marine corps casualties from World War I through the Vietnam conflict. *Military Medicine, 146*, 776-779.

Hoiberg, A. (1980). Sex and occupational differences in hospitalization rates among navy enlisted personnel. *Journal of Occupational Medicine, 22*(10), 685-690.

Hoiberg, A., & Blood, C. (1983). Age-specific morbidity among navy pilots. *Aviation, Space, and Environmental Medicine, 54*(10), 912-918.

Hoiberg, A., & Blood, C. (1985). Age-specific morbidity and mortality rates among U.S. Navy enlisted divers and controls. *Undersea Biomedical Research, 12*(2), 191-203.

Hoiberg, A., & Blood, C. (1986). Health risks of diving among U.S. Navy officers. *Undersea Biomedical Research, 13*(2), 237-245.

Jaeger, J. A., White, M. R., & Show, I. T. (1996). *EPISYS (Epidemiological Interactive System) user's guide, Version 1.0 (NHRC Tech. Doc. No. 97-1A)*. San Diego, CA: Naval Health Research Center.

Nice, D. S., & Hilton, S. M. (1990). *Sex differences in health care requirements aboard U.S. Navy ships* (Technical Report 90-2). San Diego, CA: Naval Health Research Center.

Nunnally, J. C., & Bernstein. (1994). *Psychometric theory* (Third ed.). New York: McGraw-Hill.

Ostrom, C. W., Jr. (1990). *Time series analysis: regression techniques* (Second ed.). Newbury Park, CA: Sage Publications.

Palinkas, L. A., & Coben, P. (1988). Disease and non-battle injuries among U.S. marines in Vietnam. *Military Medicine, 153*(3), 150-155.

Pugh, W. M., White, M. R., & Blood, C. G. (1989). *Disease and non-battle injury rates for navy enlisted personnel during peacetime* (Technical Report 89-51). San Diego, CA: Naval Health Research Center.

Rahe, R. H., Mahan, J. L., Jr., Arthur, R. J., & Gunderson, E. K. E. (1970). The epidemiology of illness in naval environments I. Illness types, distribution, severities, and relationship to life change. *Military Medicine, 135*(4), 443-452.

Shaw, E., Hermansen, L. A., Pugh, W., & White, M. (1991). *Disease and non-battle injuries among navy and marine corps personnel during Operation Desert Shield/Desert Storm* (Technical Report 91-44). San Diego, CA: Naval Health Research Center.

Spraycar, M. (Ed.). (1995). *Stedman's medical dictionary* (26th ed.). Baltimore, MD: Williams & Wilkins.

Tansey, W. A., Wilson, J. M., & Schaefer, K. E. (1979). Analysis of health data from 10 years of polaris submarine patrols. *Undersea Biomedical Research*, 6 (Submarine Supplement), S217-S246.

US Army Center for Health Promotion and Preventive Medicine. (1995). *Medical Surveillance Monthly Report (Vol. 1, No. 1)*. Washington, D.C.: Army Medical Surveillance Activity.

US Army Center for Health Promotion and Preventive Medicine. (1996). *Medical Surveillance Monthly Report (Vol. 2, No. 4)*. Washington, D.C.: Army Medical Surveillance Activity.

US Army Center for Health Promotion and Preventive Medicine. (1997). *Medical Surveillance Monthly Report (Vol. 3, No. 3)*. Washington, D.C.: Army Medical Surveillance Activity.

US Army Center for Health Promotion and Preventive Medicine. (1998). *Medical Surveillance Monthly Report (Vol. 4, No. 3)*. Washington, D.C.: Army Medical Surveillance Activity.

US Army Center for Health Promotion and Preventive Medicine. (1999). *Medical Surveillance Monthly Report (Vol. 5, No. 3)*. Washington, D.C.: Army Medical Surveillance Activity.

US Army Center for Health Promotion and Preventive Medicine. (2000). *Medical Surveillance Monthly Report (Vol. 6, No. 4)*. Washington, D.C.: Army Medical Surveillance Activity.

US Army Center for Health Promotion and Preventive Medicine. (2001). *Medical Surveillance Monthly Report (Vol. 7, No. 4)*. Washington, D.C.: Army Medical Surveillance Activity.

US Army Center for Health Promotion and Preventive Medicine. (2002). *Medical Surveillance Monthly Report (Vol. 8, No. 2)*. Washington, D.C.: Army Medical Surveillance Activity.

US Army Center for Health Promotion and Preventive Medicine. (2003). *Medical Surveillance Monthly Report (Vol. 9, No. 3)*. Washington, D.C.: Army Medical Surveillance Activity.

US Army Center for Health Promotion and Preventive Medicine. (2004). *Medical Surveillance Monthly Report (Vol. 10, No. 2)*. Washington, D.C.: Army Medical Surveillance Activity.

Vickers, R. R., Jr., & Hervig, L. K. (1998). *Occupational physical demands and hospitalization rates in U.S. Navy personnel* (Technical Report 98-32). San Diego, CA: Naval Health Research Center.

Wherry, R. J., Sr. (1984). *Contributions to correlational analysis*. Orlando, FL: Academic Press.

Appendix A

Full Titles for ICD Categories

Infectious	Infectious and Parasitic Diseases
Neoplasm	Neoplasms
Endocrine	Endocrine, Nutritional, and Metabolic Diseases and Immunity Disorders
Blood	Diseases of the Blood and Blood-Forming Organs
Mental	Mental Disorders
Nervous	Diseases of the Nervous System and Sense Organs
Circulatory	Diseases of the Circulatory System
Respiratory	Diseases of the Respiratory System
Digestive	Diseases of the Digestive System
Genitourinary	Diseases of the Genitourinary System
Pregnancy	Complications of Pregnancy, Childbirth, and the Puerperium
Skin	Diseases of the Skin and Subcutaneous Tissue
Musculoskeletal	Diseases of the Musculoskeletal System and Connective Tissue
Congenital	Congenital Anomalies
Perinatal	Certain Conditions Originating in the Perinatal Period
Ill-defined	Symptoms, Signs, and Ill-Defined Conditions
Injury	Injury and Poisoning
Supplementary	Supplementary Classification

Appendix C: Publications Supporting Task 3 – Simulators and Curriculum Integration

There were three works that were published and a presentation that supported Task 3:

1. Mark W. Scerbo, Leonard J. Weireter, Jr., James P. Bliss, Elizabeth A. Schmidt, and Hope Hanner, "An Examination of Surgical Skill Performance under Combat Conditions Using a Mannequin-Based Simulator in a Virtual Environment," *NATO RTO Human Factors in Medicine*, St. Pete Beach, FL., Aug., 2004.
2. Mark W. Scerbo, James P. Bliss, Elizabeth A. Schmidt, Hope S. Hanner-Bailey, Leonard J. Weireter, "Assessing Surgical Skill Training Under Hazardous Conditions in a Virtual Environment," presented at *Medicine Meets Virtual Reality XIII*, Long Beach, CA, Jan., 2005.
3. C. Donald Combs and Kara Friend, "Tracking the Domain: The Medical Modeling and Simulation Database," In J.D. Westwood et. Al. (Eds.), *Medicine Meets Virtual Reality*, 13, (90-93). Amsterdam: IOS Press, 2005.
4. C. Donald Combs and Kara Friend, "The Medical Modeling and Simulation Database," presented at *Western Simulation Multiconference '05*, New Orleans, LA, Jan., 2005.

C.1 Mark W. Scerbo, Leonard J. Weireter, Jr., James P. Bliss, Elizabeth A. Schmidt, and Hope Hanner, "An Examination of Surgical Skill Performance under Combat Conditions Using a Mannequin-Based Simulator in a Virtual Environment," *NATO RTO Human Factors in Medicine*, St. Pete Beach, FL., Aug., 2004.

Presented at the NATO RTO Human Factors in Medicine, St. Pete Beach, FL., Aug., 2004.

An Examination of Surgical Skill Performance under Combat Conditions Using a Mannequin-Based Simulator in a Virtual Environment

Dr. Mark W. Scerbo[1], Dr. Leonard J. Weireter, Jr., M.D[2], Dr. James P. Bliss[1], Ms. Elizabeth A. Schmidt[1], and Ms. Hope Hanner[1]
[1]Old Dominion University, Norfolk, VA 23259 USA
[2]Eastern Virginia Medical School, Norfolk, VA 23501 USA
mscerbo@odu.edu

SUMMARY

The present study examined the performance of a surgical procedure under simulated combat conditions. Fifteen medical students were taught to perform a tube thoracostomy on a mannequin-based simulator in a traditional medical school setting under the direction of an ATLS® certified surgeon. The participants then performed the procedure in a fully immersive CAVE virtual environment running a combat simulation including gunfire, explosions, and a virtual sniper under both daylight and nighttime conditions. The results showed that completion times depended on the order of daylight and nighttime conditions with a slight disadvantage for the nighttime condition. However, the quality of the procedures performed by the students suffered in the simulation and particularly under the nighttime conditions. Further, there were nine instances in which the participants were killed by the virtual sniper before completing the procedure. Taken together, these results suggest that the surgical skills acquired by students in a traditional medical school setting may be compromised when they are called upon to perform them under hazardous conditions. Further, the findings from this study show that virtual environments can provide a safe environment for military medical personnel to train for dangerous duty.

1.0 INTRODUCTION

Simulators have been a standard component of military training for many years in a variety of contexts including aviation, ground operations, weapons training, and decision making in command and control operations. By contrast, the use of simulation technology for training medical procedures is relatively new. Although medical simulation devices have been around since the 1940s, most of them have been little more than physical models with limited functionality. However, the current breed of medical simulators is quite sophisticated and many have impressive levels of realism. In addition, the number and variety of systems commercially available has increased dramatically over the last decade (Dawson, 2002; Satava, 2001). Further, medical schools are beginning to incorporate simulation technology into training curricula as they face increasing pressure to train physicians and surgeons to higher levels of competency, in shorter periods of time, while simultaneously improving safety (Healy, 2002).

The advantages of training with simulators are well documented. They provide an environment to train specific skills in the absence of uncontrollable influences, an unlimited number of trials to acquire skills, immediate performance feedback, and an opportunity for trainees to diagnose and treat rare or infrequent conditions. Perhaps, the most important advantage is that they permit the opportunity to train under conditions that would be too dangerous in actual operational settings. Although this last area represents a standard use of simulation for training many different skills in military contexts, it has been largely overlooked in the medical arena.

The goal of the present study was to examine the performance of surgical skills in a virtual environment (VE) under simulated combat conditions. Military medical personnel who have been in war often acknowledge that

the training they receive in traditional medical schools does not always transfer to combat situations (see for example Miller, 2003). Thus, the specific purpose of this study was to determine the extent to which surgical skills acquired in a traditional medical school might be compromised in a simulated combat scenario. Toward this end, a common emergency surgical procedure, tube thoracostomy (chest tube insertion) was selected for study.

Tube thoracostomy involves decompressing the chest cavity to release a pneumothorax or hemothorax. The procedure entails making a 1 – 2 cm transverse incision in the skin near the pectoralis major muscle lateral margin, spreading the chest wall musculature, puncturing the pleural space, and guiding a tube into the opened pleural space to permit drainage.

For this study, medical students performed a thoracostomy on a mannequin-based simulator in a CAVE virtual environment under simulated combat conditions. These conditions included visual and auditory depictions of munitions fire, gunfire, and a virtual sniper who would shoot at the participants if they did not take proper cover. Thus, the battle scenario was designed to provide a heightened sense of realism in which to examine performance of the thoracostomy procedure. The participants performed the procedure under two different lighting conditions: daytime and nighttime. The two lighting conditions were included to create different levels of workload and stress within the combat scenario. In particular, the nighttime condition was included because military medical personnel might not always have control over the visibility conditions in which they must perform. It was expected that if performance were compromised under the simulated combat scenario, it would suffer more under the nighttime visibility conditions.

2.0 METHOD

2.1 Participants

Fifteen medical students from the Eastern Virginia Medical School in Norfolk, VA participated in the study. All students were in their second or third year of training and none had prior experience with the thoracostomy procedure or the simulators used in this study. The participants received $30 as compensation for their time.

2.2 The Training System

The method used to teach tube thoracostomy was based on the ATLS® course curriculum as described in the ATLS® Instructor Manual 1997 edition. The procedure was taught using the TraumaMan® system by Simulab, Inc. TraumaMan® is a mannequin-based simulator used throughout the world for surgery education and is the only simulator approved for the ATLS® Surgical Skills Practicum by the American College of Surgeons. The TraumaMan® system is the standard training device for ATLS® courses at Eastern Virginia Medical School. The simulator includes a realistic anatomical model of the neck, chest, and abdomen with replaceable tissue components and fluid reservoirs that permit instruction on six surgical procedures. Only the thoracostomy procedure was used in the present study.

2.3 Training

Students were trained in two separate sessions and worked in groups of three or four. All students received a didactic session reviewing the indications for the thoracostomy procedure, the technical aspects of the procedure, and the potential complications that could be encountered. Following this, the students were shown

how to perform the procedure on the mannequin. Critical determinants of successful performance on the task included correct topographic anatomic landmark identification as well as whether the tube entered the pleural space. Students worked in teams of two and were allowed approximately 90 minutes of practice under the supervision of an attending surgeon qualified to teach ATLS®. Each student was required to perform a successful procedure (see below) for the instructor in less than two minutes in order to move on to the simulation session.

2.4 Virtual Environment Implementation

The VE used in the study was the CAVE (CAVE Automatic Virtual Environment). The system consisted of two main computers connected through a 100-mbps network switch. An SGI ONYX 2 computer was used to display the application in the CAVE, provide the sound playback, and read the information from the tracking device. This computer used VEGA, and IRIX 6.5. An SGI O2 computer served as the main console and was used to launch the application and issue command overrides controls during the simulation. This computer used IRIX 6.5, Motif, and Buttonfly. Images were presented on three 10x10 ft walls of the CAVE with a resolution of 1024x768.

A Radio Shack electronic beam was fixed to the top of the boxes (approximately 3 ft. above the ground). The electronic beam was used to engage the virtual sniper (see below).

2.5 Combat Simulation

The combat simulation depicted a small town under fire. One building was in flames, but most of the other combat cues were auditory in nature. Combat was simulated using the VEGA special effects module to trigger visual and auditory explosion events as well as background gunfire at specific times. The events were timed to repeat at specific intervals. The entire scenario was run in a continuous loop until the participant finished the procedure.

Day and nighttime conditions were created by adjusting the luminance intensity of the image with the time-of-day feature in the VEGA software. Under the daytime conditions, there was enough ambient illumination emanating from the walls of the CAVE to make the barricade, mannequin, and instruments easily visible. Under the nighttime conditions, however, there was very little illumination provided by the CAVE walls. Thus, the participants performed the procedure in near total darkness except for the occasional explosions that provided temporary increases in illumination.

The audio track was created using Sound Forge software. Sound samples from unrestricted sources on the internet were downloaded and filtered. Voice samples were saved in mono at a 22.1 kHz sampling rate. Background and other supplemental audio sounds included gunfire, explosions, machine gun fire, and some M1 tank fire. The files were converted to Audio Interchange File Format Version C (.AIFFC) for final presentation in the CAVE environment.

The audio files were presented over two channels. The left and right speakers were placed at approximately 225 and 315 degrees, respectively. The speakers were mounted on speaker stands at an elevation of approximately five feet. None of the audio sounds exceeded 90dB during the session.

A virtual sniper was included in the combat scenario as well. If the participant disrupted the electronic beam, an audio file would be played that provided either a warning or informed the participant that they had been killed.

2.6 General Procedure

Participants reported to the CAVE facility within two hours of training. All participants were scheduled in 20-min increments and were run individually. Participants were told they were going to play the role of an Army medic with a team of soldiers under fire. A member of the team had been injured and required a thoracostomy. Their goal was to get to the patient and perform the procedure to save his life.

They were handed a kit that contained a knife, chest tube, and clamps and were escorted into the CAVE. They were told that they would perform the procedure twice: once under daylight and once under nighttime conditions. Each attempt began with the participant standing at a starting point marked with tape on the floor. They were instructed to listen for a call for a medic. As soon as they heard the call, they were to get to the patient and perform the procedure as quickly as possible. They were not required to assess the need for the procedure. Further, they were not required to anesthetize the patient or suture/tape the tube to the patient after placing it in the pleural space. When they finished, they were told to return to the starting mark on the floor. Figure 1 shows the configuration of the CAVE facility and a participant performing the procedure.

Figure 1: Participant performing the procedure under daylight combat conditions.

The participants were also told that there was a sniper in one of the nearby buildings and that they had to take cover behind the barricade. Thus, the participants needed to perform the procedure while kneeling or lying on their stomachs. Further, if the sniper got them in his sights he would shoot to kill and they would hear a loud rifle shot. If the sniper missed, they would hear someone say "Get down." If they were hit, they would hear the phrase, "Hasta la vista, baby." At that point, they were considered dead; however, they were instructed to continue and finish the procedure. They were not fired upon again. In actuality, all participants received one warning shot if they disrupted the electronic beam. If they disrupted the beam a second time, they would be killed.

The participants each performed two sessions and the order of day and nighttime conditions was counterbalanced across participants. After the first attempt, the simulation was stopped, the mannequin was

rotated 180 degrees, and the participant performed the subsequent procedure on the opposite side of the mannequin. After their second attempt, the participants were escorted out of the CAVE and asked to complete a brief survey. They were then debriefed and allowed to offer comments or ask questions about the study. During this interval, the surgeon who had conducted the initial training session examined the mannequin and rated the participant's performance based on the criteria listed below.

2.7 Dependent Measures

There were two dependent measures: completion time and performance ratings. The total time to complete the procedure was recorded from the initial call for the medic until the participant returned to the starting mark. The performance ratings were based on three criteria: topographical location for the skin incision, tube placement in the pleural space, and posterior angulation to ensure the tube was in a dependent portion of the thorax. Ratings for each were divided into three categories: good, adequate, and poor. The following was used to assess performance:

GOOD: tube inserted at the nipple line, 4th intercostal space, between anterior and posterior axillary line, with 45 degree posterior angulation. Tube entered in the pleural space.

ADEQUATE: tube placement was within 1 cm of the criteria specified above without angulation. Tube entered the pleural space.

POOR: tube outside above 1 cm radius without angulation and/or did not enter the pleural space

3.0 RESULTS

3.1 Performance

The mean completion times for each attempt and the day and nighttime conditions are shown in Table 1. The completion times were analyzed with a mixed ANOVA for the factorial combination of attempt (first and second, the within-subjects variable) and order (groups 1 and 2, the between-subjects variable). The results revealed a significant main effect for attempt, $F(1,13) = 4.99$, $p<.05$, indicating a decrease in mean completion time from the first attempt (M=92.60, SD=24.73) to the second attempt (M=75.13, SD=20.46). However, neither the main effect of order nor the interaction between attempt and order was significant. A second mixed ANOVA was performed on day and night conditions (the within-subjects variable) and order (groups 1 and 2, the between-subjects variable). None of the main effects was significant; however, the interaction between day and night conditions and order was significant $F(1,13) = 2.54$, $p<.05$.

Table 1: Mean Completion Times for Attempts and Day/Night Conditions (standard deviations in parentheses).

	All Participants	Participants Who Were Not Killed
Attempt 1	92.60 (24.73)	99 (14.76)
Attempt 2	75.13 (20.46)	79.57 (17.27)
Day	77.87 (20.26)	86.29 (17.21)
Night	89.87 (26.57)	92.29 (20.48)

The nature of the interaction is shown in Figure 2 as a function of attempt. As can be seen in the figure, completion times were longer on the first attempt, but were clearly tied to the order of day and nighttime conditions. As would be expected, those who operated under nighttime conditions on the first attempt performed slowly. However, when this group performed their second attempt under daytime conditions they were much quicker. On the other hand, the participants who began with the daytime condition performed more quickly on their first attempt than the other group who began with the nighttime condition. However, when this second group moved onto their next attempt under nighttime conditions, their completion times improved, but not nearly as much as those of the other group.

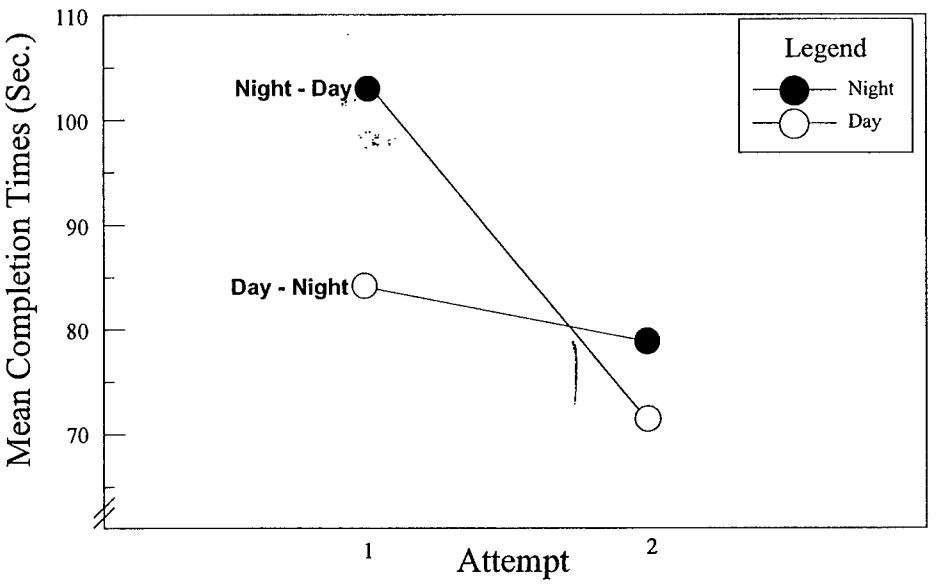


Figure 2. Mean completion times for order of day and night conditions plotted for each attempt.

The analysis of completion times included data from all participants because the sample size in this study was fairly small. However, there were several instances in which participants were "killed" by the virtual sniper. These data are shown in Table 2. As can be seen in the table, most of the participants who were killed were shot during their first attempt. Although participants in this study who were killed were allowed to continue and complete the procedure, one could argue that these data should not be included in the overall means. Thus, the mean completion times were recalculated excluding data from participants who were killed. These recalculated means are also presented in Table 1 and show that completion times increased when the scores for those who were killed are removed from the data.

Table 2: Participants "Killed" by Sniper

Condition	Attempt 1	Attempt 2
Day	4	1
Night	3	1

The performance ratings are shown in Table 3. The number of participants who received good, adequate, and poor ratings is shown for all three criteria. The ratings show that performance suffered in the simulation and particularly so for tube placement. There are no other consistent patterns among the frequencies except for a slight disadvantage for the night condition.

Table 3: Frequencies of Performance Ratings for Topographical Location, Angulation, and Tube Placement

	Location			Angulation			Tube Placement		
	Good	Adequate	Poor	Good	Adequate	Poor	Good	Adequate	Poor
Day	13	1	1	11	1	3	4	8	3
Night	11	1	3	12	0	3	3	10	2

3.2 Participant Responses

Participants completed a nine-item opinion questionnaire following the experimental session. The first item required participants to rate the realism of the thoracostomy procedure under simulated combat conditions. Although 9 participants (60%) reported that the simulation was either "somewhat" or "moderately" realistic, 40% reported that it was "extremely" or "quite" realistic. Participants were also asked to report whether the noise in the simulated combat environment distracted them. Despite the loud gunshots and explosions, 12 participants (80%) reported that the noise was not distracting or "slightly" distracting. Only 3 participants found the noise to be "moderately" or "extremely" distracting. The participants also indicated whether the lighting conditions and their physical position complicated the procedure. Twelve participants (80%) reported that the procedure was either "slightly" or "moderately" complicated under nighttime conditions, but the remaining 3 indicated that the nighttime conditions "significantly" complicated the procedure. Likewise, most of the participants reported that their physical position did not complicate the procedure. In fact, none of the participants believed that it had a "significant" or "extreme" impact on their performance.

The participants were also asked to describe any strategies they adopted for performing the procedure in the day and nighttime conditions. In the daylight condition, 57% of the respondents indicated that their main strategy was to remain low to the ground while performing the procedure. Several participants also reported that they attempted to "block out" the loud noises and focus only on the task at hand. Under nighttime conditions, approximately half of the participants (46.7%) followed the same strategy as in the daylight conditions, but the remaining participants reported that they relied more on anatomical landmarks and tactile feedback. Finally, the participants were asked to report the easiest and most difficult aspects of the experiment. Most participants agreed that the actual thoracostomy was the easiest part of the experience. On the other hand, they identified several portions of the experiment that were extremely difficult including: (1) knowing how low to remain to the ground, (2) trying to remain calm under pressure, (3) inserting the chest tube while lying down, and (4) feeling around for the equipment in the dark. In particular, participants commented that it was very difficult to remove the equipment from the bag and to place the cover back on the scalpel.

4.0 DISCUSSION

The primary goal of the present study was to examine the extent to which training in a typical medical school environment would generalize to simulated combat conditions. Students were taught how to perform a thoracostomy on the standard simulator used in ATLS® courses. In fact, they were given nearly twice as much time to practice the procedure as they would in a typical ATLS® course. By the end of the training session, each student performed the procedure to the satisfaction of the instructor, in under two minutes. They were then asked to perform the procedure in a fully immersive VE under simulated combat in daylight and nighttime conditions.

The findings were mixed. On the one hand, the results for completion times suggest that the participants were fairly efficient at performing the procedure. Overall, the mean time to complete the procedure was 84 sec (*sd* = 22) and the completion times dropped significantly from a mean of 93 sec (*sd* = 25) in the first attempt to 75 sec (*sd* = 20) in the second attempt. As expected, the completion times were affected by lighting conditions, but the results were tied to order. Specifically, the participants took 24 sec longer on average to perform the procedure under nighttime conditions if it was their first attempt, but if the nighttime conditions occurred on their second attempt, it increased completion times by only 13 sec over the daytime conditions. Thus, the participants required less time to perform the procedure on their second attempt and the effects of low visibility were less severe on their second attempt. Further, the overall completion times were not dramatically different from what they achieved at the end of their training session. Although the findings for completion times were encouraging, the results for the quality of performance were less so.

An analysis of the performance ratings showed that the ability of most medical students to perform the procedure was compromised in the simulation. Only 23% of the tube placements were judged as good and only one placement was judged good under nighttime conditions. Seventeen percent of the tubes were poorly placed. In addition, the topographical placement was judged to be poor on 13% of the attempts and the angle of placement was poor on 20% of the attempts. Thus, even though participants were able to achieve completion times in the simulation comparable to those from training, they did so by sacrificing the quality of their performance.

It is important to understand that these results present an optimistic picture of performance. There are several reasons for this. First, the results must be viewed within the context of the combat scenario. There were 9 instances where the participants failed to heed the warning shot and were "killed" before they could complete the procedure. A finding such as this is indeed troubling because it suggests a potential loss of critical medical personnel in addition to jeopardizing the safety of the patient. Further, the results in Table 1 clearly show that if the data from participants who were killed are excluded from the means, the completion times for the remaining participants are noticeably poorer.

Second, the participants performed the procedure within two hours of their initial training session. Under traditional medical school training paradigms, months or years could elapse before a student or resident would have the opportunity to perform the procedure on a genuine patient. Thus, the levels of performance obtained in this study likely represent a "best case" scenario. One might expect the performance levels seen here to become progressively worse as the interval between training and initial attempt increases.

Last, the results show that the ability to perform a newly acquired emergency surgical procedure is significantly degraded even under *simulated* combat conditions. It is quite likely that the performance problems observed in this study would be exacerbated under genuine combat conditions.

Subjective reports indicated that most participants felt the experience was fairly realistic. Several students commented that they had to make a conscious effort to remain calm and focus their attention on the task.

Although not every participant found the nighttime visibility conditions to be problematic, those who did attempted to perform the procedure by relying on tactile information. Further, some students commented that they found it challenging to perform the procedure while lying down. Collectively, these comments suggest that the students took the experience seriously and recognized the value of training outside of traditional environments.

One criticism of the present study may lie with the choice of thoracostomy as the procedure of interest. One could argue that it is unlikely this procedure would be performed in the field. That is, the injured patient normally would be moved to a safer location before performing the procedure. However, it is important to remember that the primary goal of this study was examine how skills acquired in a traditional medical school setting would hold up under stressful conditions simulated in a VE. Toward that end, we chose thoracostomy as a representative emergency procedure. Moreover, even though standard practice might dictate moving the patient to a safer environment before performing the procedure, it does not preclude the possibility that transporting the patient would be unfeasible in some situations. Thus, emergency medical personnel might be called upon to perform such a procedure to prolong a patient's life until he or she could be moved at later time.

5.0 CONCLUSION

The present study was designed to examine how the ability to perform a surgical procedure would be affected in an immersive VE. The results showed that performance was significantly degraded under the simulated combat conditions. Moreover, the levels of performance seen here are probably better than what would be expected under more realistic conditions.

To our knowledge, the present study represents the first time that performance with a standard mannequin-based medical simulator has been studied within a fully immersive VE. From this perspective, our results show that VEs can be a valuable tool for medical training because they provide a rich context in which to examine performance. The benefits of this approach are numerous. First, VEs provide a safe environment for training medical personnel on a wide range of scenarios under a variety of stressful conditions. It is no longer necessary to rely on the reports of medics or corpsmen who have been to war as the sole source of data concerning the adequacy of their training and level of preparedness for practicing medicine in combat zones. It is now possible to address specific medical training needs before personnel are deployed.

Second, VEs extend the range of applications for current medical simulators. For example, the TraumaMan® system used in this study was designed primarily as an emergency medicine training device. However, we have shown that the simulator can also be used as research tool to study performance. Obviously, other mannequin-based or even VR simulators can be used in a similar fashion.

Last, VEs provide a safe environment for studying performance under simulated hazardous conditions. Virtual environments open up the possibility of examining a wider variety of medical procedures performed under an unlimited number of conditions. More important, however, they offer a laboratory in which to study new training techniques and countermeasures for medical personnel who must perform in dangerous situations.

6.0 ACKNOWLEDGEMENT

This study was a collaborative project between the Virginia Modeling, Analysis and Simulation Center (VMASC) at Old Dominion University and the Eastern Virginia Medical School. Funding for this study was provided in part by the Naval Health Research Center through NAVAIR Orlando TSD under contract N61339-03-C-0157, entitled "The National Center for Collaboration in Medical Modeling and Simulation". The ideas and opinions presented in this paper represent the views of the authors and do not necessarily represent the views of the Department of Defense.

7.0 REFERENCES

Dawson, S. L. (2002). A critical approach to medical simulation. *Bulletin of the American College of Surgeons, 87*(11), 12-18.

Healy, G. B. (2002). The College should be instrumental in adapting simulators to education. *Bulletin of the American College of Surgeons, 87*(11), 10-11.

Miller, R. (2003, Aug.). 75th Ranger: Casualty response lessons learned. Paper presented at the Advanced Technology Applications for Combat Casualty Care Annual Meeting, St. Pete Beach, FL.

Satava, R. M. (2001). Accomplishments and challenges of surgical simulation: Dawning of the next-generation surgical education. *Surgical Endoscopy, 15,* 232-241.

C.2 Mark W. Scerbo, James P. Bliss, Elizabeth A. Schmidt, Hope S. Hanner-Bailey, Leonard J. Weireter, "Assessing Surgical Skill Training Under Hazardous Conditions in a Virtual Environment," presented at *Medicine Meets Virtual Reality XIII*, Long Beach, CA, Jan., 2005.

To be presented at Medicine Meets Virtual Reality XIII, Long Beach, CA, Jan., 2005.

Assessing Surgical Skill Training Under Hazardous Conditions in a Virtual Environment

Mark W. SCERBO, James P. BLISS,
Elizabeth A. SCHMIDT, Hope S. HANNER-BAILEY
Old Dominion University
Leonard J. WEIRETER
Eastern Virginia Medical School

Abstract. The present study examined the performance of a surgical procedure under simulated combat conditions. Eleven residents performed a cricothyroidotomy on a mannequin-based simulator in a fully immersive virtual environment running a combat simulation with a virtual sniper under both day and night time lighting conditions. The results showed that completion times improved between the first and second attempt and that differences between day and night time conditions were minimal. However, three participants were killed by the virtual sniper before completing the procedure. These results suggest that some participants' ability to allocate attention to the task and their surroundings was inappropriate even under simulated hazardous conditions. Further, this study shows that virtual environments offer the chance to study a wider variety of medical procedures performed under an unlimited number of conditions.

1. Background

Simulators have been a standard component of military training for many years in a variety of contexts including aviation, dismounted infantry operations, weapons training, and command and control operations. By contrast, using simulation technology for medical procedures is relatively new [1]. In the last five years, however, there has been a dramatic increase in the number and variety of medical simulation systems commercially available [2, 3]. Further, medical schools are now incorporating this technology into training curricula due to increased pressure to train physicians and surgeons to higher levels of competency, in less time, while simultaneously improving safety [4].

At present, most current medical simulators are designed to train basic skills or specific procedures. However, another important advantage of simulation technology is that it enables one to train under conditions that would be too dangerous in actual operational settings. Although this represents a standard use of simulation for training many different skills in military contexts, it has been largely overlooked in medicine.

Accordingly, the goal of the present study was to examine the performance of surgical skills in a virtual environment (VE) in which the operational context was simulated. Military medical personnel who have been in war have commented that traditional medical school training and practice in standard hospital settings do not always transfer to combat situations [5]. Thus, the specific purpose of this study was to determine whether surgical skills acquired in a traditional medical school and practiced in a standard hospital environment might be compromised in a simulated combat scenario.

The procedure selected for investigation in the present study was cricothyroidotomy, used when endotracheal intubation is not possible. It requires the use of one hand to lock thyroid cartilage in place and the other to make an incision in the cricoid membrane and insert tubing into the airway. The procedure was performed on a

mannequin-based simulator in a VE under simulated combat conditions including visual and auditory depictions of munitions fire, gunfire, and a virtual sniper who would shoot at the participants if they did not take proper cover. The battle scenario was designed to provide a heightened sense of realism in which to examine performance. Further, the procedure was performed under two different lighting conditions: daytime and nighttime. The two lighting conditions were included to create different levels of workload and stress within the combat scenario. In particular, the nighttime condition was included because military medical personnel might not always have control over the visibility conditions in which they must perform. It was expected that if performance were compromised under the simulated combat scenario, it would suffer more under the nighttime visibility conditions.

2. Methodology

2.1 Participants

Participants were 11 surgical residents (4 PGY-3, 3 PGY-4, 2 PGY-5, and 2 PGY-6) from Eastern Virginia Medical School in Norfolk, VA. They ranged in age from 21 to 38 years (M = 30.9, SD = 1.8). All participants had experience with cricothyroidotomy, tracheostomy, or percutaneous tracheostomy procedures. The mean reported frequency of having performed one or more of these procedures was 9 (SD = 3) and 87% of the participants indicated that they had performed one of these procedures within the last 6 months. The residents were paid $15 for their participation.

2.2 Procedural training simulator

The procedure was performed on the Simulab® Inc., TraumaMan® System. This is a mannequin-based simulator used throughout the world for surgery education and is the only simulator approved for the ATLS® Surgical Skills Practicum by the American College of Surgeons. The simulator includes a realistic anatomical model of the neck, chest, and abdomen with replaceable tissue components and fluid reservoirs that permit instruction on six surgical procedures including cricothyroidotomy.

2.3 Virtual Environment Implementation

A CAVE Automatic Virtual Environment was used to present the combat scenario. The system consisted of two main computers connected through a 100-mbps network switch. An SGI® ONYX® 2 computer was used to display the application in the CAVE, provide the sound playback, and read the information from the tracking device. This computer used MultiGen-Paradigm's Vega software running on the IRIX® 6.5 operating system. An SGI® O2 computer served as the main console and was used to launch the application and issue command overrides controls during the simulation. Images were presented on three 10x10 ft walls of the CAVE with a resolution of 1024x768. In addition, a Radio Shack electronic beam was fixed to the top of the boxes (approximately 3 ft. above the ground) to engage the virtual sniper (see below).

2.4 Combat Simulation

The combat simulation depicted a small town under fire. Combat was simulated using the Vega special effects module to trigger visual and auditory explosion events as well as

background gunfire at specific times. The events were timed to repeat at specific intervals. The scenario was run in a continuous loop until the participant was finished.

Day and nighttime conditions were created by adjusting the luminance intensity of the image with the time-of-day feature in the Vega software. Under the daytime conditions, there was enough ambient illumination emanating from the walls of the CAVE to make the barricade, mannequin, and instruments easily visible. Under the nighttime conditions, however, there was very little illumination provided by the CAVE walls. Thus, the participants performed the procedure in near total darkness except for the occasional explosions that provided temporary increases in illumination.

The audio track was created using sound samples from unrestricted sources on the Internet. They were downloaded and filtered. Voice samples were saved in monophonic format at a 22.1 kHz sampling rate. Background and other supplemental audio sounds included gunfire, explosions, machine gun fire, and some M1 tank fire. The audio files were presented over two channels. The left and right speakers were placed at approximately 225 and 315 degrees from center, respectively. The speakers were mounted on speaker stands at an elevation of approximately five feet. None of the audio sounds exceeded 90dB during the session.

A virtual sniper was included in the combat scenario as well. If the participant disrupted the electronic beam, an audio file would be played that provided either a warning or informed the participant that they had been killed.

2.5 General Procedure

All participants had formal training and experience with airway management; therefore, they were given no additional information or training for the procedure prior to the experimental session. The participants were tested individually and told they were going to play the role of an Army medic with a team of soldiers under fire. A member of the team had been injured and required a cricothyroidotomy. (The participants actually performed a cricothyroidotomy and a chest tube thoracostomy; however, only the cricothyroidotomy procedure is presented here.) Their goal was to get to the patient and perform the procedure to save his life. They were given a standard Special Operations medic kit that contained a knife and cric tube and were escorted into the CAVE. They were told that they would perform the procedure twice: once under daylight and once under nighttime conditions. Each attempt began with the participant standing at a starting point marked with tape on the floor. They were instructed to listen for a call for a medic. As soon as they heard the call, they were to get to the patient and perform the procedure as quickly as possible. They were not required to assess the need for the procedure. Further, they were not required to anesthetize the patient or secure the tube after placing it in the neck. When they finished, they were told to return to the starting mark on the floor. Figure 1 shows the configuration of the CAVE facility and a participant performing the procedure.

The participants were also told that there was a sniper in one of the nearby buildings and that they had to take cover behind the barricade. Further, if the sniper acquired a clear line of sight he would shoot to kill and they would hear a loud rifle shot. If the sniper missed, they would hear someone tell them to "Get down." If they were hit, they would hear the phrase, "Hasta la vista, baby." At that point, they were considered dead; however, they were instructed to continue and finish the procedure. They were not fired upon again. In actuality, all participants received one warning shot if they disrupted the electronic beam. If they disrupted the beam a second time, they were killed.

The order of day and nighttime conditions was counterbalanced across participants. After the first attempt, the simulation was stopped, a second mannequin was placed in the CAVE, and the participant performed the procedure again. After their second attempt, the

participants were escorted out of the CAVE and asked to complete a brief survey. During this interval, a surgeon qualified to teach ATLS® examined the mannequin and determined whether the cric tube had been placed correctly.

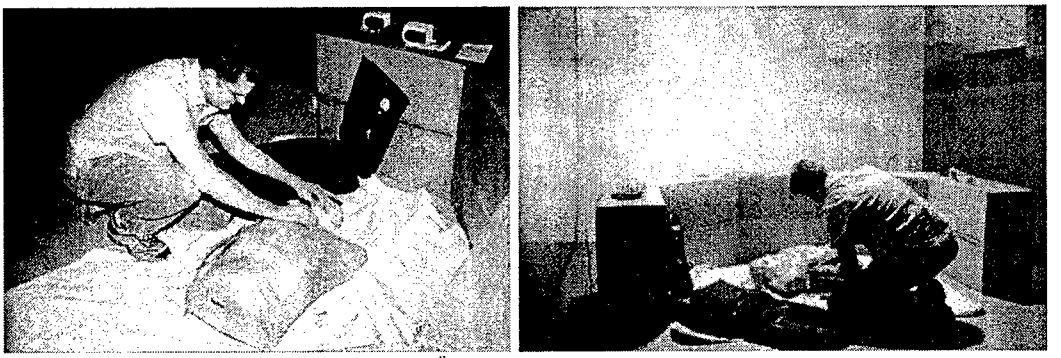


Figure 1: Participant performing the procedure under daylight combat conditions.

2.6 Dependent Measures

There were two dependent measures: completion time and performance ratings. The total time to complete the procedure was recorded from the initial call for the medic until the participant returned to the starting mark. The performance rating was based on correct placement of the tube in the trachea.

3. Results

Only one participant placed the tube incorrectly. This occurred on the first attempt under daylight conditions. Thus, completion time was the more sensitive measure.

The mean completion times for each attempt and the day and nighttime conditions are shown in Table 1. A comparison of the means showed a decrease in completion time from the first to the second attempt and this difference was statistically significant, $t(10) = 2.4$, $p<.025$. Also, although the procedure took longer to perform under night as compared to day conditions, this difference did not reach significance ($p > .05$).

The completion time analyses included data from all participants; however, there were several instances in which participants were "killed" by the virtual sniper (see Table 2). As can be seen in the table, most of the participants who were killed were shot during their first attempt. Although participants who were killed were allowed to complete the

Table 1: Mean Completion Times (in seconds) for Attempts and Day/Night Conditions (standard deviations in parentheses).

	All Participants (n = 11)	Surviving Participants (n = 8)
Attempt 1	64.9 (34.5)	60.5 (30.1)
Attempt 2	37.0 (11.2)	40.4 (10.4)
Day	48.2 (29.3)	40.2 (9.2)
Night	53.7 (28.8)	60.6 (30.4)

procedure, one could argue that these data should not be included in the overall means. Thus, the mean completion times were recalculated excluding data from participants who were killed. The means for the 8 "surviving" participants are also presented in Table 1. The overall pattern for attempts and lighting conditions remained the same; however, the difference between day and night conditions was more pronounced and approached significance ($p < .08$).

Table 2: Participants "Killed" by Sniper

Condition	
Order	
Attempt 1	3
Attempt 2	1
Lighting	
Day	3
Night	1

4. Discussion

The primary goal of the present study was to examine the performance of a surgical procedure under simulated combat conditions. Overall, the results were encouraging. Only one tube placement was judged unsatisfactory. The completion time data showed that on average, participants were able to get to the patient and perform the procedure in under a minute. Completion times were only slightly longer under the nighttime conditions, but this difference was not statistically significant. Thus, impoverished lighting conditions did not hamper performance. This finding was contrary to initial expectations, but may be due to studying a procedure that draws so heavily on the sense of touch.

The results also showed that completion times became approximately 40% quicker from the first to the second attempt irrespective of lighting conditions. This improvement in performance likely reflects increased familiarity with the testing conditions and efforts by the participants to adjust their actions to meet the task requirements. For example, in the post-experimental surveys, most participants indicated that they made a concerted effort to better organize their equipment in the medic kit so that they did not waste time fumbling around on their second attempt.

On the surface, these results seem encouraging; however, they may paint an overly optimistic picture of performance. First, there were 3 instances in which participants failed to heed the warning shot and were "killed" before they could complete the procedure. Anecdotal comments offered by some of the residents indicated that they saw little difference between the "chaos" in an Emergency Room and the "chaos" in our combat simulation and that they were able to "tune out" the noise and focus on the procedure. A finding such as this is troubling because it suggests the presence of *attentional narrowing*. Research has shown that under stressful conditions, even well trained individuals can fixate on a particular stimulus or strategy to the exclusion of other potentially relevant information [6]. Thus, even with the levels of stress created by our *simulated* combat conditions, some participants exhibited inappropriate and potentially dangerous behavior that would likely be exacerbated under genuine combat conditions.

One criticism of the present study may lie with the basic scenario. One could argue that it is unlikely this procedure would be performed under the conditions we created in our simulation. That is, the injured patient normally would be moved to a safer location before performing the procedure. Although that may be true, our primary goal was to examine how skills practiced in standard hospital settings would hold up under stressful conditions simulated in a VE. Further, even though standard practice might dictate moving the patient to a safer environment before performing the procedure, it does not preclude the possibility that transporting the patient would be unfeasible in some situations. Thus, emergency medical personnel might be called upon to perform such a procedure to prolong a patient's life until he or she could be moved at later time.

5. Conclusion

From a general perspective, the present study shows that VEs can be a valuable tool for medical training because they provide a rich context under which performance can be examined. They also provide a safe environment for training medical personnel on a wide range of scenarios under a variety of stressful conditions. The context chosen for the present study was a combat environment; however, other scenarios could also be developed addressing emergency response to natural or man made disasters or even performance within a standard hospital emergency room.

Second, VEs extend the range of applications for current medical simulators. For example, the TraumaMan® system used in this study was designed primarily as an emergency medicine training device. Obviously, other mannequin-based or even VR simulators can be used in a similar fashion. Thus, we have shown that the ability to combine varieties of simulation technology can broaden the scope of applications for medical simulation well beyond classroom instruction. Most important, however, they offer a laboratory in which to study new training techniques and countermeasures for medical personnel who must perform in dangerous situations.

Acknowledgements

This study was a collaborative project between the Virginia Modeling, Analysis and Simulation Center (VMASC) at Old Dominion University and the Eastern Virginia Medical School. Funding for this study was provided in part by the Naval Health Research Center through NAVAIR Orlando TSD under contract N61339-03-C-0157, entitled "The National Center for Collaboration in Medical Modeling and Simulation". The ideas and opinions presented in this paper represent the views of the authors and do not necessarily represent the views of the Department of Defense.

References

[1] Satava, R. M. (1993). Virtual reality surgical simulator: The first steps. *Surgical Endoscopy, 7*, 203-205.

[2] Satava, R. M. (2001). Accomplishments and challenges of surgical simulation: Dawning of the next-generation surgical education. *Surgical Endoscopy, 15*, 232-241.

[3] Dawson, S. L. (2002). A critical approach to medical simulation. *Bulletin of the American College of Surgeons, 87*(11), 12-18.

[4] Healy, G. B. (2002). The College should be instrumental in adapting simulators to education. *Bulletin of the American College of Surgeons, 87*(11), 10-11.

[5] Miller, R. (2003, Aug.). 75th Ranger: Casualty response lessons learned. Paper presented at the Advanced Technology Applications for Combat Casualty Care Annual Meeting, St. Pete Beach, FL.

[6] Wickens, C.D., & Hollands, J.G. (2000). *Engineering psychology and human performance, 3rd Ed.*, Upper Saddle River, NJ: Prentice Hall.

C.3 C. Donald Combs and Kara Friend, "Tracking the Domain: The Medical Modeling and Simulation Database," In J.D. Westwood et. al. (Eds.), *Medicine Meets Virtual Reality*, 13, (90-93). Amsterdam: IOS Press, 2005.

Medicine Meets Virtual Reality 13
James D. Westwood et al. (Eds.)
IOS Press, 2005

Tracking the Domain: The Medical Modeling and Simulation Database

C. Donald COMBS, Ph.D. [a] and Kara FRIEND, M.S. [b]
[a] *Founding Co-Director, National Center for Collaboration in Medical Modeling and Simulation Vice President for Planning and Programming Development, Eastern Virginia Medical School*
e-mail: combscd@evms.edu
[b] *Research Associate, National Center for Collaboration in Medical Modeling and Simulation Eastern Virginia Medical School*
e-mail: friendke@evms.edu

Abstract. To foster awareness of the magnitude and breadth of activity and to foster collaboration among the participants, the National Center for Collaboration in Medical Modeling and Simulation (NCCMMS) has created the Medical Modeling and Simulation Database (MMSD). The MMSD consists of two web-based, searchable compilations: one, the Research Database, that contains bibliographic information on published articles and abstracts (where available) and a second, the Companies and Projects Database, that maintains contact information for research centers, development and application programs, journals and conferences. NCCMMS is developing the MMSD to increase awareness of the breadth of the medical modeling domain and to provide a means of fostering collaboration and bringing like-minded organizations and researchers into more frequent contact with each other, thus speeding advancement of medical modelling and simulation.

The terms medical modeling and simulation suggest a myriad of ideas to various people but few people have a definitive idea of what medical modeling and simulation entails. For the purposes of this article medical modeling and simulation is defined as any form of computer or internet-based modeling that is intended to be useful in medical research or training. The field of medical modeling and simulation is rapidly expanding and the data produced by individuals and companies will soon be overwhelming to the casual researcher, and perhaps to those actively working in the field.

Medical modeling and simulation has been the subject of countless studies and articles. These articles, however, are currently spread throughout various sources on the Internet or in print. Similarly, there are hundreds of companies actively researching and creating new modeling and simulation software. Unfortunately, information on these companies is fragmented as well. If researchers could easily contact other researchers working in their field or if scientists knew where to find articles describing the efficacy of specific simulators, new research could be streamlined. To that end, the National Center for Collaboration in Medical Modeling and Simulation (NCCMMS) created the Medical Modeling and Simulation Database (MMSD).

The prototype database was created in ISI ResearchSoft's EndNote (v. 7) [1]. This is a bibliographic database specifically designed for compiling, editing and searching

Table 1. The following words were searched in combination with virtual reality, virtual medicine, computers and medicine or computer simulation.

Surgery	Computer-aided	Education	Virtual Environments
Laparoscopy	Instruction	Remote Medicine	Computer Assisted
Endoscopy	Sound	Training Validation	Surgery
Telemedicine	Data Display	Telepresence Medicine	Virtual Interface
Uteroscopy	Human Factors	Virtual Worlds	Technology
Computer Graphics	User Computer	Surgical Simulation	Surgical
Medical Applications	Interaction	Haptics	Planning

Table 2. The following search engines were used.

Pub Med (NLM)	Dogpile [1]
HealthStar (Ovid)	Medline (Ovid)
Library of Congress	Various sites will be searched using the search engine associated with the endnote software [4].

textual information. It is directly linked to several online libraries, including PubMed and OVID. The program allows the user to search an online database using keywords, dates or author names and download the relevant bibliographic data into the EndNote program. This program then sorts the entries based on the criteria specified by the user. Unfortunately this program was difficult to load on to the Internet. The database was converted to ISI ResearchSoft's Reference Manager (v. 11) as this program has a sub-program that facilitates simple web-publishing [2]. The database is currently located on a dedicated server that has been loaded with the appropriate security measures and is currently available at virtualmedicine.evms.edu.

The initial searches to fill the MMSD were designed to identify as many relevant articles as possible. Available databases were queried through the EndNote or Reference Manager programs and additional searches were performed using several internet search engines. The list of keywords used and websites queried are shown in Tables 1 and 2 respectively. After the search results were loaded into the program the researchers began the arduous task of sorting through the articles to remove any unrelated articles.

When the NCCMMS began this research, several existing catalogs of virtual reality in medicine were discovered. These sites had not been updated since 1999 (at the most recent); however, their data was incorporated into the MMSD for historical purposes. To accommodate various users' interests the database was divided into two parts; the Research Article Database (containing bibliographic data on published articles) and the Companies and Projects Database (containing information on public, private and military research as well as conferences and proceedings).

The Research Article Database contains over 13,000 entries ranging from editorials to reports on scientific experiments. The articles span almost 40 years, although almost 70% of the articles were written from 1995 to the present. The articles are arranged by author, but they can be searched in a variety of ways including journal, year, page number or keyword. Additionally, there are approximately 100 full text articles included in the database. The full text section is available in either Microsoft Word or Adobe Acrobat format. This database can be used for a variety of purposes, including surveying the domain, determining the efficacy of certain simulators, or to determine new developments in one's field of interest.

The Companies and Projects Database contains more than 600 entries detailing information about medical modeling research. This section includes conference informa-

Table 3. Virtual Medicine websites.

NAS facility at NASA Ames Research Center. http://www.nas.nasa.gov/Groups/VisTech/visWeblets.html#Commercial
Waterworth, J. A. (1999). Virtual Reality in Medicine: A Survey of the State of the Art. http://www.informatik.umu.se/~jwworth/medpage.html
Emerson, T, Prothero, J, Weghorst, S. (1994). Medicine and Virtual Reality: A Guide to the Literature (MedVR) HITL Technical Report No.B-94-1 http://www.hitl.washington.edu/projects/knowledge_base/medvr/medvr.html

tion as well as links to various research organizations. Some of these organizations are companies specializing in software or virtual reality technology while others are governmentally sponsored medical research. Also included in this section is information on magazines and journals dedicated to virtual reality in medicine. This section is invaluable to those searching for simulators or looking for information about past or planned conferences.

The MMSD will be updated monthly to prevent its obsolescence. The NCCMMS will search the Internet regularly for new publications and new companies. Once the website is fully established there will be links for researchers and developers to contact the site manager to request the inclusion of their article. The MMSD will also accept any information provided by private sector companies, government agencies or research centers. The NCCMMS will maintain the MMSD as an honest broker and will not provide one company more coverage than another except as warranted by the product lines. Not all the products listed on the site have been used or tested; rather, their listing is to provide information to the public for evaluation. The MMSD is designed to help researchers learn what is going on at other research centers and to foster collaboration between centers. This site is designed to be used by researchers, companies and the general public. The default query will search both databases to ensure the best results.

The NCCMMS will continually update the database to guarantee that the most relevant articles are available. Additional abstracts, full text articles or links to existing entries will be added as they become available. The projects section will also be updated on a regular basis including checking the integrity of the links as well as adding information to each company's website. As the database comes online, companies should start volunteering information and writing their own summaries for inclusion. This will not only ensure an accurate appraisal of the company's focus but also help make sure all relevant companies are included in the database.

The NCCMMS hopes that its 14,000 entry database will help to stimulate the field of medical modeling and simulation by allowing better collaboration and communication. While the program is still in its infancy, structural support is already in place to allow the burgeoning field to continue to develop and the MMSD to grow with it. The NCCMMS seeks through this project to ease the growing pains of the medical modeling and simulation field by creating a public forum for discussion and free exchange of ideas. For research centers, the MMSD provides a single location to search through past experiences, and for companies it provides free advertising and a new way to reach thousands of interested clients. The NCCMMS' database is filling the void created by the rapid, fragmented growth of the medical modeling and simulation field by allowing for collaboration and simplifying research.

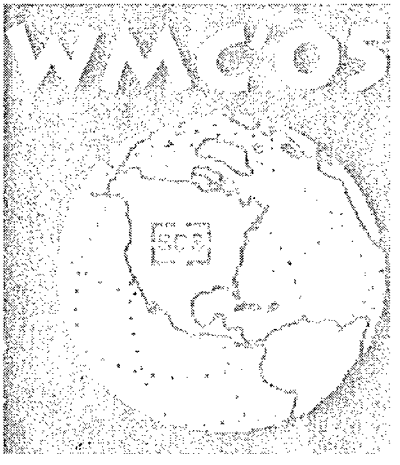
Acknowledgement

This study was a collaborative project between the Virginia Modeling, Analysis and Simulation Center (VMASC) at the Eastern Virginia Medical School and Old Dominion University. Funding for this study was provided in part by the Naval Health Research Center through NAVAIR Orlando TSD under contract N61339-03-C-0157, entitled "The National Center for Collaboration in Medical Modeling and Simulation". The ideas and opinions presented in this paper represent the views of the authors and do not necessarily represent the views of the Department of Defense.

References

[1] For more information about EndNote (v.7), see www.endnote.com.
[2] For more information about Reference Manager (v. 11), see www.refman.com.
[3] Per www.dogpile.com "Dogpile uses innovative metasearch technology to search the Internet's top search engines, including Google, Yahoo, Ask Jeeves, About, Teoma, FindWhat, LookSmart, and many more." Accessed on July 14, 2004.
[4] NCCMMS website, see nccmms.evms.edu. The database is available directly at virtualmedicine.evms.edu.

C.4 C. Donald Combs and Kara Friend, "The Medical Modeling and Simulation Database," presented at *Western Simulation Multiconference '05*, New Orleans, LA, Jan., 2005.



The Medical Modeling and Simulation Database

Journal:	*Western Simulation Multiconference*
Manuscript ID:	WMC-2005-HSS-002
Topic Area Conference:	International Conference on Health Sciences Simulation
Submission Type:	Full Paper for Review
Date Submitted by the Author:	05-Aug-2004
Complete List of Authors:	Combs, C. Donald; National Center for Collaboration in Medical Modeling and Simulation, Founding Co-Director; Eastern Virginia Medical School, Vice President for Planning and Program Development Friend, Kara; National Center for Collaboration in Medical Modeling and Simulation, Research Associate
Keywords:	simulation, database, data distribution, collaboration, research, analysis

The Medical Modeling and Simulation Database

C. Donald Combs, Ph. D.
Founding Co-Director, National Center for Collaboration in Medical Modeling and Simulation
Vice President for Planning and Programming Development,
Eastern Virginia Medical School.
combscd@evms.edu

Kara Friend, M. S.,
Research Associate, National Center for Collaboration in Medical Modeling and Simulation
Eastern Virginia Medical School
friendke@evms.edu

Keywords: database, collaboration, analysis, data distribution, research

ABSTRACT

The field of medical modeling is fragmented and isolated. Research centers frequently repeat one another's experiments and subsequently spend enormous amounts of money creating redundant software. To that end, the National Center for Collaboration in Medical Modeling and Simulation (NCCMMS) was created. Among NCCMMS' goals is one to become a meeting point for those involved in the medical modeling and simulation field. One of its first projects is the creation of Internet-based, searchable database that surveys the domain. To facilitate specific searches the Medical Modeling and Simulation Database (MMSD) has two databases; research articles and research centers/projects, which can be queried individually or in tandem. The Research Database is the larger of the two, containing over 13,000 entries at the time of publication. It contains bibliographic data about each article as well as abstracts, links to the full text or the full text (where available). The Companies and Projects database contains approximately 600 entries detailing government and private sector research and development activities as well as related conference information. This article describes the MMSD in detail and analyzes trends in such areas as research topics, primary authors, and the variety of organizations.

BACKGROUND

The terms medical modeling and simulation suggest a myriad of ideas to various people but few people have a definitive idea of what medical modeling and simulation entail. For the purposes of this article medical modeling and simulation is defined as any form of computer or internet-based modeling that is useful in medical research or training. This includes pharmacological research, virtual reality, biomedical modeling as well as procedural training and medical and health professions education. There have been thousands of articles written on the subject but these articles are currently spread throughout various sources on the Internet or in print. Similarly, there are hundreds of companies actively researching and creating new modeling and simulation software. Unfortunately, information on these companies is fragmented as well. If researchers could easily contact other researchers working in their field or if scientists knew where to find articles describing the efficacy of specific simulators, their research could be streamlined and made more effective. To that end, the National Center for Collaboration in Medical Modeling and Simulation (NCCMMS) created the MMSD. National Center for Collaboration in Medical Modeling and Simulation Eastern Virginia Medical School and Old Dominion University have received a contract from the Department of Defense (DoD) to further the development and evaluation of modeling and simulation technologies for teaching military and civilian physicians and other health professionals. When fully implemented, the award will allow the NCCMMS to conduct research in medical modeling and simulation that will strengthen clinical skills teaching centers and expand the array of simulations available to students, residents, and community practitioners. The NCCMMS has created the Medical Modeling and Simulation Database to increase awareness of the breadth of the medical modeling domain and to provide a means of fostering collaboration and bringing like-minded organizations into more frequent contact with each other.

SOFTWARE/METHODS

The prototype database was created in ISI ResearchSoft's EndNote (v. 7) [1]. This is a bibliographic database specifically designed for compiling, editing and searching textual information. It is directly linked to several online libraries, including PubMed and OVID. The program allows the user to search an online database using keywords, dates or author names and download the relevant bibliographic data into the EndNote program. This program then

sorts the entries based on the criteria specified by the user. Because this program was difficult to load on to the Internet, the database was converted to ISI ResearchSoft's Reference Manager (v. 11) [2]. Reference Manager has a subprogram that facilitates simple web-publishing. The NCCMMS has dedicated a server with the appropriate security measures to house the database, which is currently available at virtualmedicine.evms.edu.

Research

The searches to build the MMSD were designed to identify as many relevant articles as possible. Available databases were queried through the ResearchSoft program and additional searches were performed using several internet search engines. The list of keywords used and websites queried are available in Tables 1 and 2 respectively.

Table 1. The following words were searched in combination with virtual reality, virtual medicine, computers and medicine or computer simulation.

Surgery	Laparoscopy
Endoscopy	Telemedicine
Uteroscopy	Computer Graphics
Haptics	Computer-aided Instruction
Sound	Data Display
Human Factors	User Computer Interaction
Education	Remote Medicine
Training Validation	Telepresence Medicine
Virtual Worlds	Surgical Simulation
Virtual Environments	Computer Assisted Surgery
Virtual Interface Technology	Surgical Planning
Medical Applications	

Table 2. The following search engines were used:

Pub Med (NLM)	Dogpile [1]
HealthStar (Ovid)	Medline (Ovid)
Library of Congress	Various sites will be searched using the search engine associated with the endnote software. [4].

When the NCCMMS began this research, several existing catalogs of virtual reality in medicine were discovered. These sites were obsolete; however, their data was incorporated into the MMSD for historical purposes. These textual listings provided a useful starting point for the NCCMMS' researchers as well as providing a wealth of information about some of the older articles not currently accessible through the major search engines. A list of the three databases can be found in Table 3.

Table 3. Virtual Medicine websites

NAS facility at NASA Ames Research Center. http://www.nas.nasa.gov/Groups/VisTech/visWeblets.html#Commercial
Waterworth, J. A. (1999). Virtual Reality in Medicine: A Survey of the State of the Art. http://www.informatik.umu.se/~jwworth/medpage.html
Emerson, T, Prothero, J, Weghorst, S. (1994). Medicine and Virtual Reality: A Guide to the Literature (MedVR) HITL Technical Report No.B-94-1 http://www.hitl.washington.edu/projects/knowledge_base/medvr/medvr.html

Research Article Database

The Research Article Database contains over 13,000 entries ranging from editorials to reports on scientific experiments. The articles span almost 40 years, although almost 70% of the articles were written from 1995 to the present. The articles are arranged by author but they can be searched in a variety of ways including journal, year, page numbers or keywords. All articles are as complete as the sites that provided the data allowed. Some articles did not have an available abstract or translation, while some articles had links to complete text. Additionally, there are approximately 100 full text articles included in the database. The full text section is available in either Microsoft Word or Adobe Acrobat format. This database can be used for a variety of purposes, including surveying the domain, determining the efficacy of certain simulators, or to determine new developments in one's field of interest.

Companies and Projects Database

The Companies and Projects Database contains almost 700 entries detailing information about medical modeling research that is not a published article. This section includes conference information. It has links to conference websites and published proceedings where available. Additionally, this section contains links to various research organizations. Some of these organizations are companies specializing in software or virtual reality technology while others are governmentally sponsored medical research. Also included in this section is information on magazines and journals dedicated to virtual reality in medicine. There are no specific articles listings but rather general information about subscription to these journals. This section is invaluable to those searching for simulators or looking for information about past or planned conferences.

Updates

The MMSD will be updated weekly to prevent its obsolescence. Researchers will search the Internet regularly for new publications and new companies. Once the website is fully established there will be links for researchers and developers to contact the site manager to request the inclusion of their article. Emailed corrections or additions will be made by the researchers as they are received. The MMSD will also accept any information provided by private sector companies, government agencies or research centers. The NCCMMS will maintain the MMSD as an honest broker and will not provide one company more coverage than another except as warranted by the product lines. The NCCMMS does not make any claims about the efficacy of any of the products. The researchers have not tested all the products listed on the site but are merely providing information to the public for evaluation.

USES

The MMSD is designed to help researchers know what is going on at other research centers and foster collaboration between centers. This site is designed to be used by researchers, companies and the general public. The default query will search both databases to ensure the best results. If a researcher only wanted to know about companies involved in a certain type of research or only wanted research articles, the site allows the user to query a single database. The advanced search also further refines the results the user can obtain. Reference Manager also allows the user to download their search results as a text file or into an ISI ResearchSoft program. There are a number of other features available through Reference Manager and EndNote that are unfortunately not available to the general public through the NCCMMS website. Some of the more useful features allow researchers to survey the domain by looking at the most prolific writers (either in general or for specific years) and the most frequently used keywords. Additional ISI ResearchSoft programs will allow researchers to graphically analyze the data contained in the MMSD.

FUTURE DIRECTIONS

The NCCMMS will be continually updating the database to ensure that the most relevant articles are available. Additionally, researchers will be adding abstracts, full text articles or links to existing entries as they become available. The projects section will also be updated on a regular basis including checking the integrity of the links as well as adding information to each company's website. Researchers are hoping that companies will start volunteering information and writing their own summaries for inclusion. This will not only ensure an accurate appraisal of the company's focus but also help make sure all relevant companies are included in the database. This does not mean that companies who do not submit information will be excluded as researchers will continue to perform periodic Internet searches to ensure accuracy. The NCCMMS is also hoping to foster collaborative efforts by providing a complete listing of companies. Each company has a website and contact information if it was available. Companies that have changed names or been renamed can be searched by their old name or their new one. NCCMMS plans to develop a website with message boards and news items to facilitate the sharing of research and information. Researchers hope that companies and research centers will take advantage of the information they have made available to further simulation research as a whole. Conference information (including registration and location) will be made available through the MMSD as it becomes available.

SUMMARY

The NCCMMS has created the MMSD to foster collaboration between medical modeling in all its forms. It is the hope of the researchers that a nonbiased database will allow researcher centers to determine the state of the research field. For example, a company trying to build a virtual histology lab might work with a company creating a virtual human for a more complete educational tool. The researchers at NCCMMS hope that their database can create a more efficient research field through the elimination of redundant research as well as streamlining the entire virtual medical research process. There are admittedly a few shortcomings in the prototype version of the MMSD, which researchers are currently working to remedy. There are bound to be articles or companies that the researchers have missed. If something is missing a quick email will remedy the situation. The researchers will also continue to search on their own to find missed articles or companies. The researchers will periodically scan the database to ensure the relevance of the articles included therein. The database has also recently been moved to its own dedicated server to ensure faster searches as well as better security.

The NCCMMS hopes that its 14,000 entry database will help to stimulate the field of medical modeling and simulation by allowing better collaboration and communication. While the program is still in its infancy, structural support is already in place to allow the burgeoning field to continue to develop and the MMSD to grow with it. The NCCMMS researchers hope to ease the growing pains of this field by creating a public forum for discussion and free exchange of ideas. For the research centers, the MMSD promises a single location to search through past experiences, and for the companies it promises free advertising and a new way to reach thousands of interested clients.

The NCCMMS' database is filling the void created by the rapid fragmented growth of the medical modeling and simulation field by allowing for collaboration and simplifying research.

REFERENCES

[1] For more information about EndNote (v.7), see www.endnote.com
[2] For more information about Reference Manager (v. 11), see www.refman.com
[3] Per www.dogpile.com "Dogpile uses innovative metasearch technology to search the Internet's top search engines, including Google, Yahoo, Ask Jeeves, About, Teoma, FindWhat, LookSmart, and many more." Accessed on July 14, 2004.
[4] NCCMMS website, see virtualmedicine.evms.edu

ACKNOWLEDGEMENT

This study was a collaborative project between the Virginia Modeling, Analysis and Simulation Center (VMASC) at Old Dominion University and the Eastern Virginia Medical School. Funding for this study was provided in part by the Naval Health Research Center through NAVAIR Orlando TSD under contract N61339-03-C-0157, entitled "The National Center for Collaboration in Medical Modeling and Simulation". The ideas and opinions presented in this paper represent the views of the authors and do not necessarily represent the views of the Department of Defense.

Appendix D: Publication Supporting Task 4 – Regional Medical Response Simulations

D.1 John A. Sokolowski, Suchitra Manepalli, Giridhar Manepalli, and Matt Davis, "Designing an Agent-Based Population Model to Support Mass Casualty Planning," Virginia Modeling, Analysis and Simulation Center, Old Dominion University, 2004.

1. John A. Sokolowski, Suchitra Manepalli, Giridhar Manepalli, and Matt Davis, "Designing an Agent-Based Population Model to Support Mass Casualty Planning," Virginia Modeling, Analysis and Simulation Center, Old Dominion University, 2004.

Designing an Agent-Based Population Model to Support Mass Casualty Planning

John A. Sokolowski, Ph.D., Suchitra Manepalli, Giridhar Manepalli, Matt Davis
Virginia Modeling, Analysis & Simulation Center
Old Dominion University
Norfolk, VA 23529
jsokolow@odu.edu

Keywords: multi-agent, population, mass casualty.

1. INTRODUCTION

There is considerable utility in having a model capable of simulating the population of a city or region. A few uses include investigating the spread of a disease, analyzing social interactions, land use planning, and designing a health care system. This paper sets forth the conceptual design for such a population model that will support the training and analysis requirements of hospital administrators, public health officials, and city emergency managers in the area of mass casualty planning and management.

The United States has not had a mass casualty event that involved more than about three thousand victims since the early 1900s [1]. Yet, with the threat of terrorists bringing weapons of mass destruction to the U. S. shores, planning for such an attack is a critical requirement of emergency managers. The release of biological or radiological agents can pose a significant health threat to the people living in the area of the release. The implications of such releases in a given area must be well understood. A population model that simulates the response to these events can help provide insight into the issues public health officials must face and help them formulate policy and decisions that minimize the impact of such events.

The remainder of this paper describes the conceptual design of the population model underlying and supporting mass casualty simulation and analysis. It begins with a description of the design requirements for the population model and then explains how these requirements were fulfilled.

2. CONCEPTUAL MODEL DESIGN

A conceptual model design is the first step in realizing the implementation of a simulation system within a computational environment. It provides the basis for which code can be written to capture the desired level of model fidelity. This section describes the conceptual elements making up the population model design.

2.1 Conceptual Design Requirements

A key piece to developing a mass casualty model is the modeling of the population response to such events. Specific population modeling requirements were devised to ensure population behavior was captured at the appropriate level to achieve the overall goals of meeting the training and analysis requirements of health managers. The following is a list of those requirements.

1. **Model population behavior at the individual person level.** To help understand how the aggregate population as a whole will behave, it is necessary to define behaviors at the individual level and then observe how each of these individual behaviors contributes to the aggregate response. This concept is analogous to swarm models where individual entities are provided with simple behavioral rules that produce very complex group behaviors [2]. The same is true for population behavior. For example, half dozen patients with specific symptoms showing up at local health care facilities may be enough to alert the public health care system of a potential epidemic or at least an abnormal pattern of sickness.
2. **Accurately portray regional population demographics.** Demographics play a significant role in the overall behavior of a population. Older populations may behave in one way when faced with a certain event while a younger population may proceed in an entirely different manner. The same goes for women vs. men and one ethnic background compared to another. The model should be capable of replicating these demographic differences and capturing their different behavioral patterns.
3. **Portray the daily and weekly routines of individuals.** In general, individuals tend to follow a daily and weekly routine, which could include

going to work or school, participating in recreational activities, and gathering in public places such as theaters or stadiums. Additionally, these routines change depending on the day of the week and holidays. These routines provide the opportunity for a variety of person-to-person interactions and determine where a person is physically located at any given time of day. All of these have a bearing on an outcome of a mass casualty event. For example, an attack on a school on the weekend would not have nearly the impact as an attack during a normal school day.

4. **Portray the social structure of individuals.** By their very nature, people become part of a specific social structure that is unique to the individual. Examples of social links include family, classmates, coworkers, and neighbors. This social structure forms a complex network of individuals and links, which have a direct bearing on how individuals behave and interact. Social structure is significant because relationships can manifest different effects within the population. For example, if a person with a well-connected social network becomes infected with a deadly disease, he or she has a higher probability of passing that disease on to numerous individuals compared to a person with a limited social network.
5. **Use an object-oriented design.** An object-oriented approach facilitates encoding the characteristics and behaviors of the individual population entities in a concise manner.
6. **Provide for model scalability.** Design the system to operate in a distributed and parallel processing environment to facilitate scaling the population to one representative of a large metropolitan area. Scalability is necessary to support the computations required for populations of over one million entities to ensure the model is useful in complex scenarios with potentially thousands of casualties.
7. **Model the effects of various types of mass casualty events on the health and behavior of individuals in a population.** The model must account for injury, sickness, and mortality and the associated population behaviors that would manifest themselves during events that would cause mass casualties.

2.2 Conceptual Design Details

The previous section described the design requirements. This section explains how these requirements have been incorporated into the conceptual design.

The design decision that had the most influence on the overall model structure was that of using a multiagent system approach for representing the population. Weiss defines a multiagent system as a system in which several interacting, intelligent agents pursue some set of goals or perform some set of tasks [3]. This description reflects the general characteristics of how a population behaves in that individuals interact with one another on a daily basis depending on their social structure while pursuing the goals that are the most important to them at the time. The agents that make up a multiagent system are autonomous, i.e. they make decisions on their own based on their goals and perception of their environment. This characteristic is conducive to modeling individual members of a population who have freedom of choice on what decisions to make.

Once a multiagent design approach was adopted, work began on designing the necessary agent behaviors that would reflect an accurate aggregate population behavior. A finite state machine (FSM) was the methodology chosen to represent and implement agent behavior. One can describe a FSM as an abstract machine that defines a finite set of conditions of existence (called "states"), a set of behaviors or actions performed in each of those states, and a set of events that cause changes in states according to a finite and well-defined rule set [4]. FSMs are an efficient way to specify constraints of the overall behavior of a system. Being in a state means that the system responds only to a subset of all allowed inputs, produces only a subset of possible responses, and changes state directly to only a subset of all possible states.

Agents must be in the same or nearby locations to experience physical interactions such as the passing of a disease. In addition, external physical events such as explosions or earthquakes affect agents in the same location in the same manner. Thus, location was chosen as the main state variable for each population agent. An agent's FSM is centered on deciding where the agent should go next. The location decision is influenced by the role an agent plays in the population, the agent's health status, and knowledge of its external environment. For this model, there are three possible roles for an agent: adult, student, or child. An agent's health status is a reflection of the degree of wellness or sickness of the agent. Knowledge of the external environment includes time of day, day of week, and health status of relatives. Time of day and day of week influence what location an agent is striving to reach because each agent follows a daily routine based on its role. For example, most adult agents will strive to go to work during the weekday. Student agents will tend to go to school. Regional demographics determine the specific distribution of locations each agent may attain. Health status of relatives also influences an agent's location decision. If a child is ill,

Table 1. Bayesian Belief Table

Role	Health	Time	Day	P(work)	P(home)	P(school)	P(pharmacy)	P(doctor)	P(hospital)	P(recreation)
Adult	Well	Day	Weekday	.7	.2	0	0	0	0	.1
Adult	Well	Day	Weekend	.05	.45	0	0	0	0	.5

one or both of its parent agents may decide to take the child agent to the doctor or the hospital depending on the seriousness of the sickness or injury.

Location decisions are made via a Bayesian Belief Network (BBN). BBN is a data structure that represents the dependencies among variables and gives a concise description of any full joint probability distribution [5]. It is an efficient way to describe a complete probability distribution that governs an agent's location decision. Agent goals are implicit in the belief network. The possible choices for an agent's location are: home, work, school, pharmacy, doctor, hospital, store, stadium, and theater. Multiple locations for each of these exist based on area demographics. Agents will proceed to the work, home, and school locations to which they are assigned. Potential pharmacy, doctor, hospital, store, stadium, and theater location decisions are based on the nearest available facility. Table 1 is a partial representation of the BBN. Each row has probabilities that sum to one because an agent must decide on a location for every set of conditions. For brevity, not all locations and their associated probabilities are shown.

As pointed out earlier, location facilitates interaction among agents. No agent interactions occur if a pair of agents is not collocated. If they are collocated, then their social relationship governs the interaction. Social relationships are defined by a social network structure that describes agents and their relationships. Relationships can be such things as family, coworker, classmate, or neighbor. Since relationships define specific ties between two agents, they govern the probability of an interaction. If no specific social relationship exits, then the probability of interaction between two agents is small and random. This type of interaction accounts for a chance encounter between two agents in the same location with no social network connection.

Each agent has a data structure that defines its social network. For the purpose of this model, the social network describes relationships between pairs of agents. For example, an agent may have a social network that consists of ten members and their corresponding relationships. Member 1 might be a child. Member 2 might be a coworker. Member 3 might be a neighbor. The social network provides the basis for the interactions that can occur on a periodic basis. The diversity of each agent's social network and the stochastic nature of location decisions provides for a complex response surface that can account for a wide variety of outcomes characteristic of human behavior.

As previously noted, time of day affects an agent's location decision. Time is divided up into three periods, day (0800-1600), evening (1600-2400), and night (2400-0800). Each of these three time periods are divided in half accounting for a total of six time steps for the simulation in a one day period of time. The number of time steps can be adjusted as necessary if agent interactions must be determined more often such as in a rapidly occurring disaster like an earthquake.

The population model is being developed in the Java programming language. Implementing it in Java allows the population simulation to run on multiple hardware platforms with varying operating systems. Additionally, using Java assists in model scalability through the JavaSpaces service. JavaSpaces is a core service provided by Sun Microsystems Jini technology [6]. Javaspaces allows a developer to bring independent software processes together into what appears to be a single distributed application. Since each agent in the population model is autonomous and acts independently of the other agents, it can be processed in parallel with other agents with inter agent coordination accomplished through message passing from one agent to another. JavaSpaces allows independent processors in a cluster or distributed computing environment to process waiting agents as processors become available. Once a processor is through with one agent, it can process another waiting agent thus minimizing processor idle time. This parallel processing capability allows the number of agents to scale up in size in a near linear fashion as the number of processors is increased. Testing on a 64 processor Linux cluster has shown the capability to process 100,000 agents with one year's worth of interactions in about 36 hours of processing time.

3. CONCLUSION

Work continues on refining and testing the population model design. Daily agent movement and interactions have been validated by comparing them to demographic patterns. Incorporation of agent behaviors for various mass casualty events is in progress.

4. REFERENCES

[1] Influenza 1918 "...the worst epidemic in United States History", http://www.pbs.org/wgbh/amex/influenza/, 11-30-2004.

[2] Tarasewich, P. and McMullen, P. R., 2002, "Swarm Intelligence: Power in Numbers", Communications of the ACM, 45, no.8, 62-67.

[3] Weiss, G., 1999, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence.*, The MIT Press, Cambridge MA.

[4] Gill, A., 1962, *Introduction to the Theory of Finite-State Machines.*, McGraw-Hill, New York.

[5] Russell, S. and Norvig, P., 2003, *Artificial Intelligence: A Modern Approach*, Pearson Education Inc., Upper Saddle River NJ.

[6] Jini Network Technology, http://java.sun.com/developer/products/jini/index.jsp, 11-30-2004.

5. BIOGRAPHIES

JOHN SOKOLOWSKI, PH.D. is a senior research scientist at Old Dominion University's Virginia Modeling, Analysis & Simulation Center (VMASC). He holds a Bachelor of Science degree in Computer Science from Purdue University, a Master of Engineering Management from Old Dominion University (ODU), and a PhD in Engineering with a concentration in the Modeling and Simulation from ODU. Prior to coming to VMASC, he spent 27 years in the Navy as a submarine officer with his final assignment as Head, Modeling and Simulation Division, Joint Warfighting Center, U.S. Joint Forces Command. His research interests include human behavior modeling, multi-agent system simulation, and simulation architectures supporting homeland security.

SUCHITRA MANEPALLI is a Masters student in Computer Science at Old Dominion University. She obtained her Bachelor's Degree in Computer Science from Old Dominion University (ODU) in 2002 with an honors degree. She is currently working as a research assistant for Virginia Modeling, Analysis and Simulation Center under Dr. Sokolowski. Currently, she is one of the key researchers in Trend Analysis and Detection, a Digital Library Project in collaboration with Harvard University.

GIRIDHAR MANEPALLI is a Masters Student in Computer Science at Old Dominion University. He is involved in numerous projects, which involve Artificial Intelligence and Digital Libraries. His research areas include Peer-to-Peer Information Systems, Parallel Computation, Artificial Intelligence, Digital Preservation, Knowledge Discovery, and Linear Algebra Applications in Data Mining. Currently, He is a Research Assistant for Archive Ingest and Handling Test Project. The project is in collaboration with Library of Congress, Harvard University, Stanford University, and Johns Hopkins University. His work involves integration of MPEG-21 Multimedia Framework with self-deployable Digital Objects, called Buckets for Digital Preservation. He also works in Population Behavior Simulation Agent Model at Virginia Modeling, Analysis and Simulation Center under Dr. Sokolowski.

MATT DAVIS is an undergraduate student majoring in Computer Science at Old Dominion University. His research interests include Java applications development in a distributed computer environment.

FINAL TECHNICAL REPORT

DISTRIBUTION AND DISTRIBUTION STATEMENT

Naval Health Research Center (NHRC) Technical Point of Contact
Dr. Frank Garland
Naval Health Research Center
Code 02B
PO Box 85122
San Diego, CA 92186-5122
Tel: (619) 553-8480
Fax: (619) 524-9888
E-mail: garland@nhrc.navy.mil

Naval Air Systems Command Naval Air Warfare Center (NAVAIR)
Training Systems Division (TSD) Project Manager (PJM)
NAVAIR ORLANDO TSD PJM
Attn: Ron Root, TSD 11E21
12350 Research Parkway
Orlando, FL 32826-3275
E-mail: Ronald.Root@navy.mil

Naval Air Systems Command Naval Air Warfare Center (NAVAIR)
Training Systems Division (TSD) Procuring Contracting Officer (PCO)
NAVAIR ORLANDO TSD PCO
Attn: Steven Harnig 25333 c/o
Beverly Gordon
12350 Research Parkway
Orlando, FL 32826-3275
E-mail: Beverly.Gordon@navy.mil

Office of Naval Research (ONR) Administrative Contracting Officer (ACO)
James Gandy ACO
Atlanta Regional Office
Office of Naval Research
Suite 4R15
100 Alabama Street, SW
Atlanta, GA 30303-3104
Tel: (404) 562-1602
Fax: (404) 562-1610
E-mail: gandyj@onr.navy.mil

FINAL TECHNICAL REPORT

Defense Technical Information Center (DTIC) (2 copies)
Defense Technical Information Center (DTIC)
Attn: DTIC-OC
Suite 0944
8725 John J. Kingman Road
Fort Belvoir, VA22060-6218

Approved for public release;
distribution is unlimited.